Arunachalam, V.
Quantitative Genetics From the Perspective of Mendelian and Molecular Breeding.
Crop Improvement, 28 (2001): 129-156.

Crop Improv., 28 (2): 129-56 (2001)

QUANTITATIVE GENETICS FROM THE PERSPECTIVE OF MENDELIAN AND MOLECULAR BREEDING

V. ARUNACHALAM

M. S. Swaminathan Research Foundation
Third Cross Road, Taramani institutional Area, Chennai 600 113.

Quantitative trait (QT) is a term of central importance in the fields of biology and agriculture. As the term indicates, QT refers to characters that can be measured on a quantitative scale. This is, however, a textual definition but we also need a scientific one. Such a definition would necessarily be with reference to the field or activity in which QT is to be employed. Therefore we restrict our attention to agriculture, and more specifically to crop breeding and improvement.

The QT measurements can be viewed as a scatter of numerical values and therefore dealt with using concepts of experimental statistics and inference. Indeed, such is the view of a majority of breeders and scientists working in the area of crop improvement. Short of effective training in the elements of mathematics and statistics, they consider the subject dealing with QTs as one beyond their comprehension. On the other hand, it is known that significant improvement in crop productivity can be achieved only with significant improvement in the associated component QTs. Usually courses in Plant Breeding deal with the basics of Mendelian Genetics and laws of inheritance. But the subject of Quantitative Genetics integrates the concepts of Genetics and morphometric trait improvement. Teaching and research in this subject, however, remain restricted by inadequate specialization and skill.

In the backdrop, a cross-section of scientists including geneticists have a lingering question in their mind about the discipline to which Quantitative Genetics belongs as a subject. In view of the fact that it deals with phenotypes and aims to generate, edit, modify, orient or upscale plant breeding paradigms, they wonder whether the subject belongs to the discipline of Genetics or Plant Breeding. In essence, the roots of such doubts lie in a debillitated introduction to the subject. Just as the subjects of cytogenetics, developmental genetics, cell biology and molecular biology perceive (genes and) genotypes from different angles, so does Quantitative Genetics from their phenotypic expression. Naturally therefore, Quantitative Genetics rightly belongs to the discipline of Genetics. Further this subject is unique in that it interprets phenotypic trait expression as a function of the underlying genotype, and serves as the most effective bridge between principles of Genetics and fundamentals of Plant Breeding.

Quantitative Genetics provides the foundation for breeding methodologies. As research in quantitative genetics advances, the foundation gets reinforced,

breeding methodologies refined and new methods proposed. But insufficient exposure to quantitative genetics limits the option to following mechanically the breeding methods proposed elsewhere.

Diallel selective mating initially attempted in barley, a self-pollinated crop (Jensen, 1970) was one such example. Diallel crosses starting with 7 parents were continued in F_1 and other generations obtained by selfing as needed. Simple selection procedures were suggested to derive a pure line or population. Even with 7-parent diallel, we get 21 F_1s and the method suggested selective 2-parent crosses (each parent itself is an F_1 and hence the crosses here are 4-way crosses; limit set was 21) between F_1s. Such 2-parent choice among 21 F_1s can be made in $21C_2$ ways = 210 ways. 21 crosses can then be selected in 210 C_{21} way, an astronomical figure indeed. In self-pollinating crops, the suggested 21-parent four way crosses in F_1 will be time- and cost-intensive. When an analysis of resulting data is not available and special benefits of the design are unconvincing, the very need for such methods will also be unconvincing. Yet, at that time, many breeders in India attempted this method only on self-pollinated crops following published literature.

To do justice to all developments in Quantitative Genetics in a review would be a Herculean task. Therefore the basics of Quantitative Genetics are projected bringing out the basis behind multi-trait characterization. The topics of evaluating genetic divergence and choice of parents for initiating crosses in a breeding program were chosen as examples to illustrate the fundamental differences between mendelian and Molecular breeding.

It is not uncommon to find advanced books approaching the subject of Quantitative Genetics through the pathway of statistics introducing models and estimation of parameters. Linking analytical results with basic genetic theory or its practical applications does not always get a focus. When the subject branches further into applications to plant breeding, an impression is left behind that Quantitative Genetics deals only with genetic models, gives strenuous statistical formulation associated with the breeding procedures and the computation and presentation of the associated statistics are a ritualistic must! This spells a strong need to place the basics on a different perspective.

Numerous books on Quantitative Genetics deal with various methods of breeding; quite a few of them start with genetic material governed by a single diallelic gene. In that case, the number of possible matings and therefore breeding methods are restricted. Progeny, genotypes necessarily belong to one of the three genotypic categories - dominant or the recessive homozygote and the heterozygote. The progeny mean and breeding improvement in various fillal generations can be calculated with relative ease. From such relatively simple single gene methods, transition to complex models of mating based on many genes is usually attempted. Assumptions like independent and additive genes are made to enable extension of results of a single gene model to complex situations. But further treatment of the subject masks the underlying assumptions made earlier. Most often, only the

formulae emanating from such a treatment get focused and the assumptions on which they were based remain latent.

Quantitative genetic modeling generates new breeding avenues and upscales the specifics of current breeding methods. Strategically, it provides a set of quantitative parameters for evaluating the efficiency of the methods and for inferences on the next phase of breeding. For example, in the process of estimating the average degree of dominance, the mating system in an F_2 population governed by a single diallelic gene was modeled. That provided a comparison of the efficiency of three designs of mating (Comstock and Robinson, 1948; 1952), popularly known as North Carolina Designs I, II, and III. One of them, Design II, was a precursor of the line X tester mating design, which was later given a specific treatment by Kempthorne (1957). Similarly the diallel cross initially treated on a purely theoretical plane using genotypes of a single diallelic gene (Jinks and Hayman, 1953; Jinks, 1954; Hayman, 1954) was given a broad-based, breeding-oriented treatment later (Griffing, 1956). This work described four diallel crossing methods, ranging from n^2 to n(n-1)/2 crosses (where n is the number of parents and where appropriate, parental selfs are also included). For instance, with 10 parents, diallel matings can be set up with 100, 55, 90 or 45 crosses following Griffing's four models and genetic information obtained with relative precision. Depending on the need, capacity and preference, a breeder can choose the method desired. Such a flexibility incorporated in Griffing's diallel crossing system has made it popular among breeders.

There are numerous designs of mating each of which has an underlying quantitative genetic model enabling estimation of relevant quantitative parameters. They help not only in assessing the efficiency of the mating system but also in predicting expected genetic advance. Enumerating all the breeding methods developed so far and explaining the underlying models and parameters would be tenuous. The crop under development, the breeding objective like pure line, hybrid or population, the natural pollinating system and the genetic status of initiating material (like advanced breeding lines, single or multiple cross, a diversified gene pool etc.) are all highly relevant to the process of improvement. If, all of them are taken into account, the number of cases to deal with will multiply many-fold. Therefore, this paper stops short of dealing with them.

As earlier said, basic models of Quantitative Genetics provide the foil for improved methods of breeding. Once rightly positioned, the methods of inference, limitations of inferred results and further steps or breeding become logical derivatives. It would be interesting to illustrate, though not exhaustively, a smooth transition of the basics to a fundamental breeding process.

Developments in molecular biology have given rise to molecular markers, which are numerous compared to morphological markers. Those modern developments are said to energize classical breeding methods and accelerate breeding progress through precise perception of the genetic changes. It would be useful, therefore, to understand the basics of molecular markers in the light of Mendelian genes. In

that process, the feasible and hypothetical strengths of molecular markers would also be discussed. As an example, the choice of parents for initiating a single cross will be illustrated both by classical and molecular methods.

The Basics :

1. QT and its genetic basis :

One way to characterize an individual is by using attributes that occur in one of two visual forms, for example, leaf colour (yellow or green). Such visualization is usually 'qualitative'. Mendel in his experiments with plant hybridization (with pea plant) choose to characterize individual plants on such 'either-or' attributes. All the 7 traits he chose to measure (which he called 'experiments') were of this type. They were, as we now know : 1. Form of seed (round or angular), 2. Colour of seed albumen (yellow or green), 3. Colour of seed coat (white or gray), 4. Form of ripe pods (regular or constricted), 5. Colour of unripe pods (green or yellow), 6. Position of flowers (distributed along main stem or bunched at the top), 7. Length of stem (6 to 7 feet or ¾ to 1½ feet).

Though Mendel simplified quantitative measurements into qualitative (either-or) ones, he was well aware of this oversimplification. For instance, he recognized that the seed albumen (endosperm) could be pale yellow, bright yellow-and-orange through a number of gradations to more or less intense green. To identify, as a first step, a rule of inheritance, he might have thought it safe to confine to only two highly contrasting colour shades, which would also enable unambiguous scoring of individuals in segregating generations.

In those experiments, Mendel obtained the 3:1 segregation in F_2 generation, which led to the illustrious laws of inheritance. What is relevant here was the inference that, each of the 7 traits, as measured by Mendel, was essentially controlled by a single diallelic gene. But some results of Mendel's experiments were imperfect and contentious. As sole experiments, such results are not totally unexpected. Yet a case of rare occurrence of recessive parent (1 in 31) in F_2 made him reflect on the colouring pattern of parental traits and realize that flower colour controlled by gene A could be a combination of the effects of individual alleles, A_1, A_2, A_3, ... A_n of the gene A. When any one or more of them occur in hybrid and F_2 combinations, independent colours could develop. As a consequence, the colour combinations were more than two. We now know that a single gene can have 'n' multiple alleles and diploid genotypes in that case are n (n+1)/2. A sample size of 31 could then be inadequate to trap more than one white-coloured recessive parent.

Mendel also dealt with 2 factors controlled by genes A and B. His experimental results showed that the marginal genotypic frequencies (e.g. genotypes AA - -, Aa - -, aa - -, where - - represents all combinations of the gene B, namely, BB, Bb and bb, are marginal genotypes. Similarly, -- BB, -- Bb and - - bb are the other marginal genotypes) always followed Mendelian segregation. We now know this result from the theoretical frequencies of the genotypes governed by 2 genes. We realize that,

when Mendel observed the colour of seed albumen, he did not take into account the colour of seed coat, for example. Thus for each single factor he observed only the marginal genotypes which showed single gene segregation. This facilitated the postulation of his laws of inheritance. Mendel's original paper presented in detail in Bennett (1965) would provide a worthwhile reading and should be read to understand and appreciate the logic of inference from designed and carefully conducted plant experiments.

The above exposition highlights the fact that QTs are, in general, governed by more than a single gene and genes may have more than two alleles as well. With modern techniques and knowledge, every trait can quantitatively be measured with good accuracy. Thus flower colour can now be measured as intensity of colour pigmentation, disease resistance as a combination of traits like disease infection index, % leaf area damaged, AUDPC (area under disease progress curve) and so on. In effect therefore, it would not be an exaggeration to say that almost all plant traits are essentially quantitative.

In essence, a galaxy of QTs, each of which can be governed by many genes, can describe an individual. For instance, flowering time, plant height, number of tillers per plant, length of panicle, number of grains per panicle and the like are all traits that describe individual varieties. To characterize an individual variety fully, it is essential to use a multiple trait description. Otherwise, differentiating between individuals could be deceptive. For example, two varieties may flower at the same time; the grain yield of one may be far more than the other. In that case, characterizing the varieties using the single trait, flowering time, would become deceptive, as the varieties would be ranked to be quite similar in their performance. But it would not be so, if differentiated on grain yield. The difficulty, however, is that the investigator does not know a *priori* which single trait (like grain yield, in this example) will help discriminating between varieties. Further, it is possible that varieties may be ranked similar on more than one trait too. A safe strategy would therefore be to characterize individuals on multiple traits. Naturally, multivariate analysis would be the right choice for discriminating between such individuals.

QTs defining plant varieties are spread over their entire growth period. Further, as the plant growth progresses, traits measured in those stages become increasingly complex. Thus seedling, flowering, post-flowering, seed filling and maturity define broadly the growth phases with QTs measured in them showing increasing complexity. Grain yield is a trait of major interest measured in the post-maturity phase. Logically therefore, it is a function of various traits in the earlier phases, though the relationship is neither obvious nor can be discovered with ease. Further, the relationship between yield and a host of traits measured in earlier growth phases is dynamic and influenced by a variety of factors like the crop and the growth environment at the site where it is grown. It is easy to visualize two individuals to show highly similar yields; but it is almost impossible that such similarity would exist for each component QT measured in various stages of plant growth. In general therefore, breeding initiatives use yield components in preference to *per se* yield.

2. Phenotype and its relation with Genotype :

Mendel's experiments were an initial attempt to relate QT measurements with the underlying genotype. Expressed in an alternative way, the QT (one at a time) [Phenotype] is the expressed value of the underlying genotype in the experimental environment (location, year, agro ecology etc.). Initially, it was thought that the phenotype, P results as an additive function of the genotype (G) and environment (E);

in other words, $\mathbf{P = G + E}$ (1)

Note that this relationship equates a phenotypic value (measurement) like grain yield with its genotypic value (can, at the most, be conjectured) and environmental value (cannot again be directly measured). Thus we have a peculiar equation expressing a dependent variable, P (only one capable of direct measurement) as an additive function of two independent variables, G and E (both of which do not admit of direct measurement)! It was R.A. Fisher, one of the originators of experimental designs and Quantitative Genetics who analyzed this intriguing problem. He realized that, when G and E are independent, the second-degree statistic, namely, variance is given by

$\sigma_P^2 = \sigma_G^2 + \sigma_E^2$ where σ^2 denotes variance. (2)

By growing the plant material in a field design like r.b.d. (randomized block design) and employing the analysis of variance (ANOVA) on the measured QT (P values), one can estimate σ_E^2. For example, if 10 varieties were grown in 4 replications, analysis of yield data will give the Error variation with 27 degrees of freedom. The corresponding Error mean squares will provide an estimate of σ_E^2. In Fisher's conception, Error is associated with an environment which cannot be measured; for example, error variation in the above example, could have arisen due to soil heterogeneity in the field plots, climatic conditions that prevailed during the experiment including moisture availability, soil type, temperature, incidence of various stresses and the like. We note that the overall environment here cannnot be measured by a quantitative value for each variety in each replication. Hence environmental variance cannot be estimated directly.

Environment here needs to be understood rightly. It is clearly different from a treatment whose effect can be measured. But as the subject of Quantitative Genetics progressed, it became a practice to use agronomic treatments like fertilizer doses, irrigation levels, and plant densities as equivalents of Environment. It is essential to note that they are various treatments dealt with in detail in the subject area of Agronomy and there are designs of experiments to estimate their effects and variances. Therefore, they do not qualify to be used as Environment. In general, locations and years are the most relevant environments for genetic studies.

On a close look at the equation (2) relating variances, it would appear that it subsumes many genotypes and their phenotypes. On the other hand, it would

equally apply to a number of phenotypic values of a particular genotype expressed in various environments in which case, the appropriate equation would be,

$$P_n = G + E_n \qquad (3)$$

where n = 1,2, ..., k are k environments. Based on statistical principles, if n is large (theoretically k tending to infinite number), the values of E_n which are essentially deviations of P_n from the true value of G in various environments would be expected to be positive in some environments and negative in some others. When n becomes very large, the positive and negative effects cancel each other to make the sum of such E_ns tend to zero. In other words,

$$\Sigma P_n = nG + \Sigma E_n$$

$$= nG + 0$$

Hence, $G = \Sigma P_n / n = P$

Theoretically, this is one way of obtaining the value of G provided 'n' is very large and the genotype is stable with no intra-genotype variation.

In general, ANOVA of data from r.b.d. provides an estimate of σ_E^2 (=Error mean square) and σ_G^2 (obtained from the expected mean square of 'Between varieties' source of variation). Using the equation (2), we get σ_P^2 which helps to estimate Broad sence heritability as $h_{BS}^2 = \sigma_G^2/\sigma_P^2$ (4)

It must be noted, however, that the model given by (1) is oversimplified and untenable; for, it is now known that G and E are never independent and genotype X environment interaction is universal. These facts invalidate equation (2). In practical breeding applications, therefore, broad sense h^2 quite often fails to have a meaningful value.

It was Fisher who conceived the genotypes TT and Tt as resulting from successive substitution of gene T for t in the genotype tt. It is easily seen that one substitution in tt gives rise to Tt and another substitution gives rise to TT [for details, see Fisher (1941)]. From this analogy, it was suggested that a genotypic value G can be thought of the sum of additive value (A) of genes constituting the genotype and a residual, which is actually the dominance deviation or the dominance value (D) [for more details, see Falconer and Mackay, 1996]. Initially this concept was applied to a single gene model and estimates of A and D obtained.

Thus, P = G + E

or P = A + D + E (5)

Alternatively, the genotypic values of TT, Tt and tt in a single gene case, admit of 2 degrees of freedom (among the 3 genotypes), which are unequivocally accounted for by the additive(A) and dominance (D) effects. Further, A and D are statistically independent. Since E is still an effect not estimable directly, it is

assumed to be absent to give, as a first approximation of equation (1), P = G. Under this assumption, the usual procedure is to fit a regression line, P = G = A + D, to estimate A. The deviation (P - A) would then estimate D (see, for details, Li, 1955). In accordance with the Fisherian concepts, it is also possible to *assign* a genetic value equal to the number of dominant genes carried by a genotype and obtain the estimates of A and σ_A^2 (Fisher, 1941; Arunachalam and Owen, 1971). In other words, TT is assigned a genetic value 2, Tt a value 1 and tt a value zero. Measured from the mean, these values become X_{TT}=1, X_{Tt}=0, Xtt= -1, where X denotes the *assigned* genetic values for the respective phenotypes. Equation (2) can then be written as

$$\sigma_P^2 = \sigma_G^2 + \sigma_E^2$$

$$= \sigma_A^2 + \sigma_D^2 + \sigma_E^2 \qquad (6)$$

Obviously, the theory of partitioning phenotypic variances becomes quite complex when more than one gene is considered, for, there would be a number of interaction effects involving A and D. When linkages between genes are also taken into account, the complexity would further increase (Arunachalam 1988; Arunachalam and owen, 1971).

We note that, in this process, a population TT, Tt and tt governed by a single gene T is considered, the frequencies of the genotypes under random mating are p^2, 2pq and q^2 where p is the frequency of the allele T and q that of allele t, p + q = 1.

Then the QT values of the genotypes TT, Tt and tt would be

$y_{TT} = m + d$

$y_{Tt} = m + h$

$y_{tt} = m - d$, where

m = mean value of homozygotes = ½ $(y_{TT} + y_{tt})$

d = ½ $(y_{TT} - y_{tt})$ and

h = y_{Tt} - m [see Mather and Jinks, 1977]

Then the additive effect A and dominance effect D would be

$A = d + (q - p)\, h$, also denoted by α

$D = 2h$ (7)

The additive and dominance variances would then be given by

$\sigma_A^2 = 2pq\, \alpha^2$ and

$\sigma_D^2 = 4p^2q^2\, h^2$ (8)

The above formulae further refine the estimate of σ_G^2 obtained before. This refinement leads to the narrow sense heritability,

$$h^2_{NS} = \sigma_A^2/\sigma_P^2 \qquad (9)$$

It must be noted that h^2_{NS} is always less than h^2_{BS}. In plant breeding experiments, one does find values contradicting this basic fact. One reason is that σ_A^2 is estimated through other means, like correlation between relatives, and not by (8) and fitted in the simplified model (6) leading to incompatibilities.

Another important point to bear in mind is that, to begin with, d, -d and h are merely values of the two homozygotes and heterozygote and nothing more, even if we consider random mating population. The additive effect given in (7) is a function of the d and h values and also the gene frequency. But dominance value is independent of gene frequency and twice the value of heterozygote. Only if we make a further assumption that p = q or the frequencies of T and t in the original population are equal [or alternatively, the population is the F_2 of the corss, TT X tt], then d becomes equal to the additive value ($A = \alpha = d$). This further assumption adds to the already untenable assumptions under which expressions for A and D were obtained.

If, erroneously, it is taken for granted, that additive value = d and dominance value = 2h always, it would appear that additive value as given in (7) contains dominance value and hence additive variance also contains dominance variance, as stated in Mather and Jinks (1971) [see also the correct exposition in Arunachalam, 1976].

The above well-known expressions have been explained in some detail mainly to underscore the spate of assumptions made to arrive at them. Therefore, they should be borne in mind when additive and dominance varainces are used in breeding applications. In addition, most of the molecular breeding theory leans heavily on the basic formulation detailed so far, another reason why it should be understood unambiguously.

3. Mendelian Gene and Molecular marker :

A knowledge of the basics behind Mendelian gene and Molecular marker will help a proper comprehension of the quantitative analysis based on classical genetics and modern biology. They are therefore briefly explained here.

3.1 Mendalian Gene :

In the development of fundamentals, we characterized an individual phenotype on the value of a single QT. Further the QT was modeled to be governed by a single diallelic gene giving rise to three genotypes. Genotypic frequencies in filial generations are calculated using mendel's laws of inheritance. The phenotypic values were used, under stringent assumptions, to obtain the additive genetic and dominance variances. They help in formulating a genetic basis of breeding for QT

improvement. The gene considered here is a classical gene whose expression is measured by the QT (phenotype) and which follows Mendel's laws of inheritance. Hence we call it a 'mendelian gene'.

3.2 Molecular Marker :

In contrast, the advent of molecular biology gave rise to a set of molecular markers (e.g. RFLP) with analogous properties. The salient difference is that those marker-genes lack a direct expression on their own. The logic behind molecular markers has been clearly explained in Kochert (unpublished), which is summarized below for the benefit of those who cannot access the unpublished reference.

Figure 1 represents a cross between two homozygous parents; one parent, P1 had green leaves with genotype GG and its RFLP marker pattern showed a band at the 6 kb position. Its homozygous RFLP marker state can be denoted as 6kb/6kb. Similarly, the other parent P2 had yellow leaves with genotype gg, its RFLP showed a band at the 8 kb position and the marker state was donoted as 8kb/8kb.

From Mendelian laws of inheritance, the F_1 genotype would be Gg, and would carry bands both at the 6 and 8 kb positions (see Fig. 1) with the marker state 6kb/8kb. Due to dominance, the F1 phenotype would have green leaves. In the F2, as we know, there would be two phenotypic classes C1 and C2 segregating in the ratio 3:1 with frequencies ¾ and ¼ . It is known that the phenotypic class C1 contains P1 with frequency ¼ and recombinants (Gg heterozygotes) with frequency ½ . The class C2 contains P2. Phenotypically, parent P1 and recombinants in the class C1 would all have green leaves. To detect the recombinants, a breeder has to self the plants in C1 (or resort to any other appropriate crossing program) and note the segregation which would take one more generation, But with molecular markers, the recombinants can easily be located in the class C1 itself as those with both the 6 and 8 kb bands present. The easy and clear identification of recombinant genotypes is claimed to be the unique advantage of molecular markers over the classical Mendelian methods; more significantly, molecular markers clearly define and detect genotypes while morphological markers depend on phenotypes for genotypic identification, it is further observed that, morphological markers would be disabled from identifying genotypes when environmental effect masks variation making phenotypes look alike. Only further pedigree testing can resolve the problem.

Molecular markers, based on DNA, are independent and do not have an association with environment (contrasting Mendelian genes controlling a QT) validating the model P= G as an initiating step. So it is affirmed that molecular markers provide genetic values directly. But a very tight linkage between a molecular markers provide genetic gene governing a QT is invoked to measure G as equal to P. Where such tight linkages are unavailable, it is not possible to estimate G. It must be recognized that P of a QT has an environment effect in it. Regardless of the fact that molecular markers are independent of E, when a measure of P is used

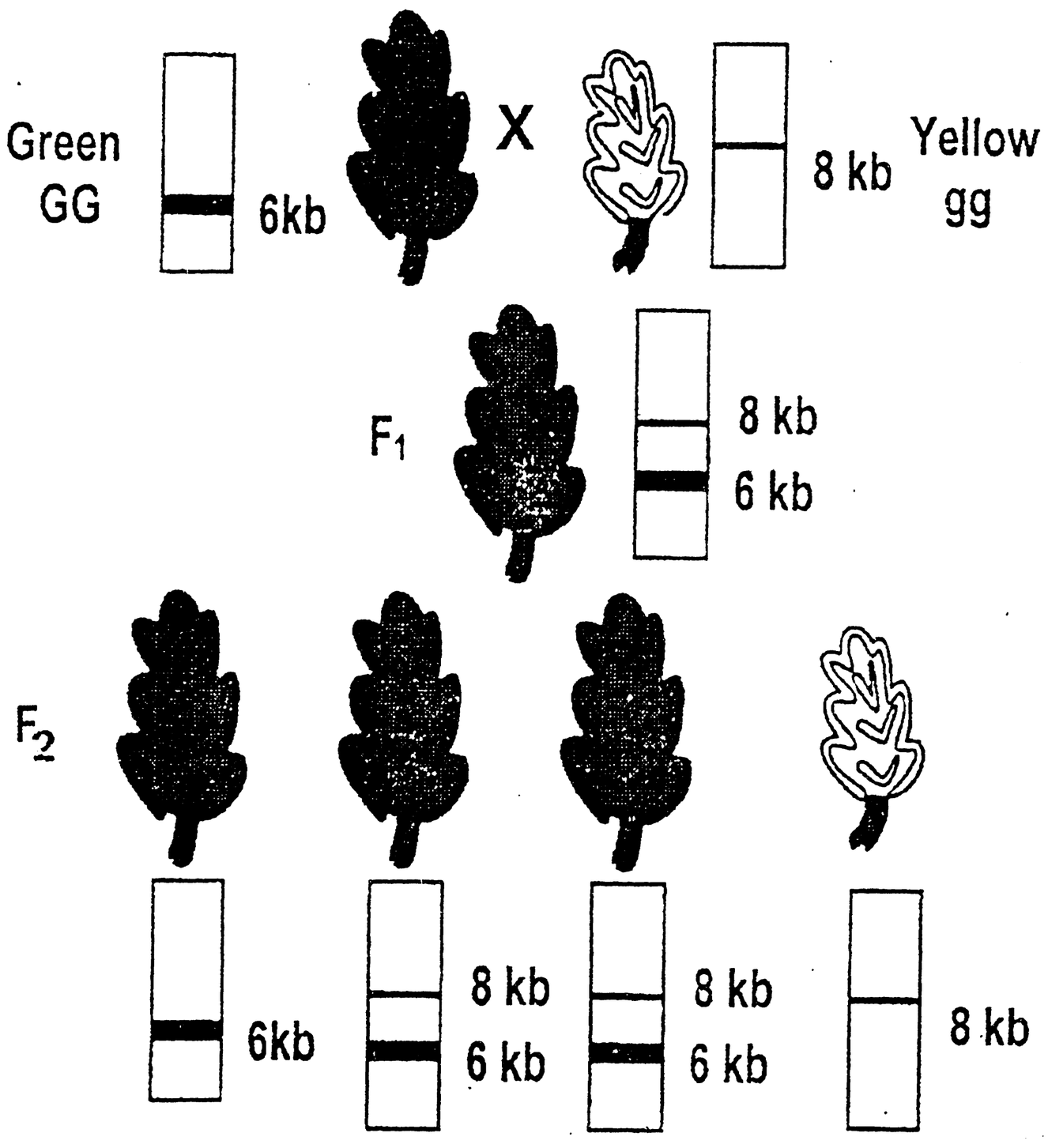


Figure 1. RFLP Inheritance Pattern (Adapted from G. Kochert)

as the value of G (P ≡ G) of a molecular genotype, E effect gets automatically injected into it.

3.3 The contrast between Mendelian Gene and Molecular marker :

At best therefore, molecular techniques recognize genotypes, TT, Tt and tt and *assign* them genetic values, 2, 1 or zero; in other words, $X_{TT} = 1$, $X_{Tt} = 0$, $X_{tt} = -1$, as mentioned earlier. Those values are fixed and remain constant for every marker genotype regardless of organisms (be it a crop plant, animal or human) or QTs (be it flowering time, seedling or plant height, or grain yield) (see, for example, Martin *et al.*, 1989; Jansen and Stam, 1994; Zhang *et al.*, 1994).

Though familiar, we need to note the succinct difference of this formulation from that of a Mendelian gene where additive genetic effect, A (= α) and dominance genetic effect D are functions of the phenotypic values of the three genotypes and their frequency.

As we have made it clear, the basic assumption P = G is a first approximation; once we deal with second degree parameters, more distinctive weight is given to σ_E^2 and the theory and inferences are set accordingly. We clearly comprehend in

this context, the inseparable nature of genetic and environmental effects producing measurable phenotypes.

The contrasting conceptualization of Mendelian gene and molecular marker diverges further when inferences based on QTs are invoked to synergise with those based on molecular marker parameters. The areas of such inferences are widely different and beyond the purview of this review. Yet, the conceptual divergence between morphometric and molecular approaches will be apparent when we consider genetic divergence, as measured by a quantitative distance parameter.

4. Differentiating between individuals :

It is most usual to regard QTs as governed by many genes, each producing a small major effect on its own but with pronounced inter-genic interaction effects (epistasis). The observed values of QTs would vary across environments (E effect described in § 2). But, the environmental effect, E, following a Gaussian error (normal) distribution, is expected to affect the genetic effect, G making it err more often around mean of its distribution than shaking it violently towards the extremes. In this light, the value of a QT can legitimately be regarded as the expression of genes governing it in an environment. The variation in QTs across genotypes (individuals) would then reflect the genetic variation confounded only by the environmental effects (specific only to that environment). Investigations aimed at a *comparative evaluation* of the genetic potential of varieties *in a specific environment* could then use QT variation (to represent the genetic variation) in this light.

4.1 Genetic Divergence :

Genetic distinction or relatedness is an important concept in plant breeding. Such distinction is made using a measure of genetic distance or genetic divergence. In essence, a set of varieties is sought to be classified into groups on their *inter-se* genetic distance. Ideally, a group should be such that the 'within-group' distance is far smaller than the 'between-group' distance. Measures of genetic divergence and methods of grouping are varied in number, application and underlying logic. We choose, for a comparative account, highly utilized multivariate divergence (D^2) and Nei's distance, popular in most of the molecular applications.

4.1.1 Multivariate divergence :

As observed earlier, an individual is better defined by a set of 'n' traits. For example, let us say that an individual is described by 2 traits only - flowering time (x_1) and number of tillers (x2). If, further, we assume that the traits x1 and x2 are independent, we can represent them in a two-dimensional graph with x1 and x2 as two rectangular axes. Then two individuals described by the traits x1 and x2 can be plotted on the graph. The distance between them could be used to decide whether they are close or highly divergent. But we know that x_1 and x_2 are not independent, in which case the individuals will be represented in oblique axes. This concept can

be extended to 'n' traits with 'n' oblique axes corresponding to them. Distance between varieties can then be calculated using the multivariate distance statistic, D^2 due to Mahalanobis. Since genes govern traits, and environmental variation can be estimated, it would be feasible to conceive D^2 as representing genetic distance [For a detailed exposition, see Arunachalam, 1981].

In breeding for improvement, individuals (genotypes) of a population are conceived to be wholly described by a set of QTs. How many QTs and what type of QTs are live questions. Over years, it has been generally accepted that those related to the fitness of an organism (Murty and Arunachalam, 1966), particularly those related to survival, deserve preference.

Visualizing the QTs to fix a variety firmly on a multivariate plane, a multidimensional distance measure (provided by Mahalanobis' D^2 statistic) is used to assess the inter-varietal genetic divergence (see, for example, Rao, 1952). When 'k' varieties (individuals) are considered, we would have k (k-1)/2 pair wise D^2 representing genetic distance or divergence between all possible pairs of varieties. Scanning the pattern of D^2 values, a raw grouping is obtained which is then fine-tuned following a method due to Tocher, explained in Rao (1952). A norm is set to decide the difference between any two D^2 (or alternatively the divergence between any two varieties) within a group that can defensibly be tolerated. That norm cannot be set following a fixed rule (as was done in Singh and Choudhary, 1983) as it is dependent on the range and *distribution* of observed D^2 values. Once the groups are firmed up, intra- and inter-group average D^2 delineate those groups clearly.

The method first developed and applied in1965 in *Brassica* (Murty *et al.*, 1965) was adopted widely across crops and varied genetic material for the next decade and half. In many instances, the grouping was confirmed by principal component analysis too. Computer software for calculating D^2- values, which otherwise involve complex computations, was also developed (Murty and Arunachalam, 1966) enabling grouping based on a large set of QTs. Yet dealing with a large number of varieties renders scanning a large number of distances (between all possible pairs of varieties) quite arduous. Therefore a modified method of grouping was developed (Vairavan *et al.*, 1973; Durga Prasad *et al.*, 1985). As the utility of D^2- statistic in genetic classification became more and more pronounced, further modifications were made (Arunachalam and Bandyopadhyay, 1984; Arunachalam *et al.*, 1998) to enable grouping of varieties into four divergence classes uniformly. This enabled an unbiased comparison across experiments (or even crops where such a process is justifiable).

However, a few questions used to linger in the minds of geneticists and breeders regarding D^2 analysis, the most frequent of which were :

* Since distances are computed using phenotypic values, D^2 does not represent genetic but only phenotypic divergence.

*Nowhere genotypic frequencies are used; in that sense, genetic inferences can hardly be derived.

The process used in computing D^2 answers the first doubt. A transformation of QT values using 'error' covariance matrix between the 'n' variates (corresponding to the 'n' traits) is used to effectively deal with genotypic variation (see also § 3). The second doubt may not be relevant when we deal with 'n' varieties, germplasm entities and the like since hardly can we know their genotypes or frequency. Further, any genetic model assumed in that context, would remain conceptual and may not concern the individual entities in any meaningful way.

It is relevant, in this context, to note that distance statistic as a measure of genetic divergence has been tested across crops, environments, germplasm, breeding material and varied instances over the past four decades in India and productive inferences derived from its targeted use remained as a landmark in crop breeding.

4.1.2 Molecular Divergence :

Refined and developing techniques are of a vast kind now to identify molecular markers spanning the entirestretch of DNA with coding and non-coding regions. Initially, molecular genetic distance was developed to study the affinity between human populations where, unlike in plants, it is not easy to locate quantitative traits that define an individual satisfactorily. Under a set of conditions, the genetic distance concept was first developed (Nei, 1972) to deal with local populations within species. It used essentially the probability that a gene from population 1 and a gene from population 2 are identical. It was argued that this distance measure, D, was the most appropriate unlike others where "it is not clear what biological unit that is being measured". This claim (Nei, 1972) is debateable in the light of the distance measure defined on QTs using D^2 statistic. At the same time, it was recognized that Nei's measure of D is affected by a number of factors such as detectability of gene differences, varying rate of nucleotide substitution at different loci etc. The applicability of the measure is extended to any population, including selfing populations, if populations can be adequately defined (though it is not clear what is really meant!). Further D depends solely on gene frequencies than on genotypic frequencies. This work was extended to measure "nucleotide diversity" under a restrictive assumption that all nucleotides are distributed at random over the DNA sequences with a given G + C content (Nei and Li, 1979). The concepts underlying genetic distance measure, D have been further refined (Nei, 1987), though the basic definition given earlier remained the same. It is now used almost exclusively in molecular genetic studies and a software incorporating it, is readily available for molecular data analysis (see, for example, Wang *et al.*, 1992).

4.2 A comparative analysis of genetic divergence measured by morphometric and molecular distances ;

The fundamental aim in measuring genetic distances is to identify individuals (essentially phenotypes, equivalently, genotypes expressing in an environment) whose performances are similar. But in view of environmental effects, phenotypic performance may be deceptive in identifying genetic affinity. At the same time identifying genetic affinity through the frequency of molecular markers (independent of environmental effects) may not be sustained where environment plays a definite role in the individual's performance. In reality, we face a different type of problem. It is known that molecular techniques can identify only single diallelic molecular markers. QTs are usually polygenic with high epistasis. Even when more number of molecular markers is identified, they are independent and free from epistatic effects. Further, in a single gene case, genotypic frequencies are determined by gene frequencies alone. But in multigenic case and with linkage, the above observation fails to hold. Recognizing that the genetic distance given by D depends solely on gene and not on genotypic frequencies (Nei, 1972), it is a moot question whether such a measure can identify genotypic affinity. The conceptual differences between molecular and morphometric distances (Table 1) would help to make the distinction more lucid.

Table 1. The salient differences between molecular (Nei's D statistic) and morphometric (Mahalanobis' D^2 statistic) distance

	Molecular distance	**Morphometric distance**
Based on	allelic concentration in the discovered molecular markers	expressed QTs in the concerned environment
Individuals defined on	molecular markers	a set of identified QTs distinguishing performance
Genetic constitution	a number of non-interacting independent markers	a number of interacting genes displaying epistatic effects
Environment	independent	highly dependent
Formula	function of probabilities like the probability of a gene from population 1 and a gene from population 2 being identical	multivariate distance statistic using environmental variation to correct phenotypic variation
Genetic classification	usually dendogram allowing flexibility of inferences	Repeatable grouping into 4 divergence classes (see § 4.3.1)

4.3 Grouping of individuals on *inter-se* genetic divergence :

A breeder would like to know the pattern of genetic divergence in the varieties intended to be used for further genetic improvement. Though the divergence between pairs of varieties can be measured using distances based on morphometric traits or molecular markers (§ 4.1.1 and § 4.1.2), the number of such distances increases sharply with the number of varieties. Therefore there is a need to group varieties on their *inter-se* genetic divergence.

4.3.1 Morphometric Grouping :

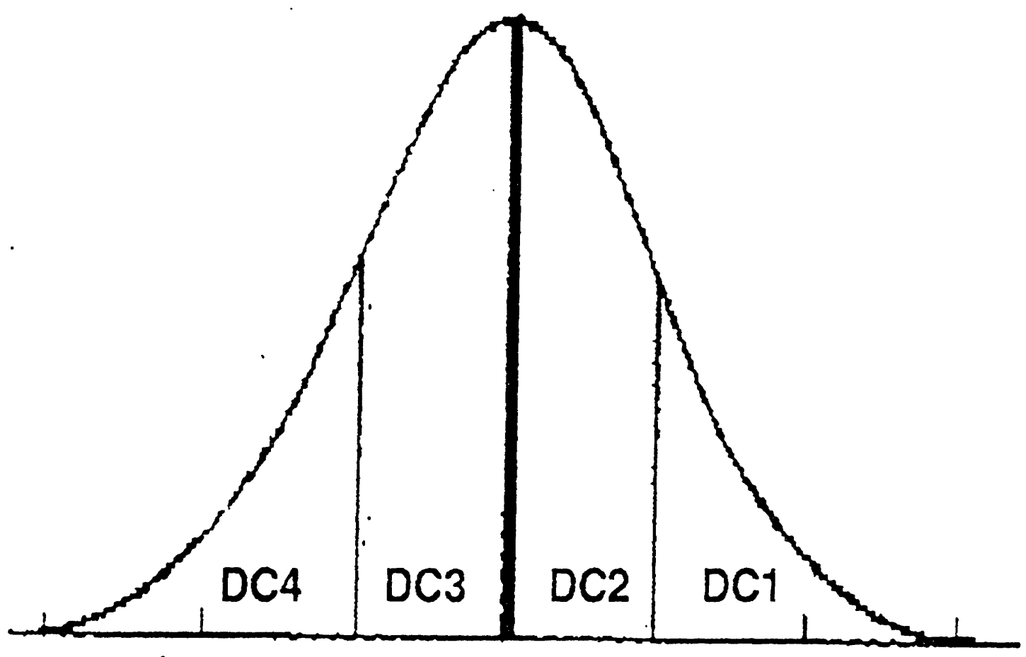
Morphometric grouping arranges the varieties into a few clearly delineated groups (§ 4.1.1). The inter-group distances will define how close or divergent the groups (and the varieties contained in them) are, as measured by the QTs. However, a number of restrictions surfaced in its application over more than two decades.

* The method of Tocher uses an arbitrary norm to be set by the experimenter to define an allowable D^2, as the difference between average D^2 and the quantity [total value of D^2/total number of D^2] as and when varieties are added to a group at the time of consolidating final grouping (see, for details, Rao, 1952). In such a procedure, different experiments with different crops conducted in different environments could give different number of groups. In addition, different investigators may set different norms in which case, the same material from the same experiment could also give different number of groups. In such a case, the inferences would suffer limitations since comparison would have to be made among varying number of groups across experiments.

* Further, while software to compute D^2s is available (Murty and Arunachalam, 1967), software to group varieties is not. This has been a handicap to research students in institutions where insturction in multivariate analytical methods is weak or absent and students seldom have access to do the grouping manually.

A new methodology was developed after years' of work with the aim of removing the limitations of Tocher's method of grouping. In short, the distribution of the D^2 values was assessed using the mean (m) and standard deviation (s). Keeping in mind the property of a standard normal distribution that 68% of the total area lies between m-s and m+s (Fig. 2), the D^2 values were set in 4 divergence classes (DC) using the following criterion :

This criterion was found to be optimal as discovered from several divergent examples from various crops (Arunachalam and Bandyopadhyay, 1984). If more number of divergence classes are desired, one can set more dividing points; for example, m-s, m- ½ s, m, m+ ½ s, m+s as dividing points would give 6 classes. However, for many of the breeding initiatives, four divergence classes were found to suffice.



DC	X
1	$h \geq X \geq m+s$
2	$m+s > + \geq m$
3	$m > X \geq m-s$
4	$m-s > X \geq k$

X : D^2 value; h : highest;
k : lowest K

Fig. 2. The logic behind divergence classes

The advantages of uniformly 4 divergence classes are :

* there are no arbitrary norms involved in grouping, this procedure would thus lead to unique groups.

* experiments repeated over locations, seasons etc. could be comparatively evaluated on uniform grouping.

* where there is scope, experiments across crops and global environments can even be comparatively evaluated for convergence of concepts; e.g. optimal divergence level for parental choice to enhance the probability of F_1 heterosis.

Grouping based on genetic divergence confers the following advantages :

* Provides insight into the level of divergence among groups of varieties.

* divergence interpreted on the nature and origin of varieties can provide leads on their genetic nature like self-incompatibility (Murty *et al.*, 1965) or their evolutionary pattern (Murty *et al.*, 1966) for example.

* provides clues to choice of parents from divergent groups to decide crosses from which breeding processes can be initiated

The last one is crucial in breeding initiatives. It was earlier established (Falconer and Mackey, 1996) that heterosis over mid-parent is a function of dominance effect and the gene frequency differences of parental populations, comprising single gene diploids. Translated to a general case, it was postulated that parents showing a high degree of genetic divergence are likely to produce heterosis. However, further theoretical research established that, too great a divergence would not ensure heterosis (Cress 1966) and additive effects and additive type of interactions alone (without even dominance) can produce heterosis in two

gene systems (Arunachalam, 1977). These findings made it clear that there is a concerted need to find limits to parental divergence for enhancing the probability of F_1 heterosis. After analyzing a number of experiments critically and taking into account the conceptual frame due to Langham (1961), the following hypothesis was set on a confirmatory plane.

When varieties are grouped based on genetic divergence (measured by D^2) into 4 divergence classes, DC1 to DC4, the chances of F_1 heterosis enhance when parents are selected from the intermediate classes, DC3 and DC2 [Arunachalam and Bandyopadhyay,1984].

It was found, in general, that the divergence class DC1 contains high values of D^2; therefore heterosis in crosses between parents from this class should be rare. At times, crosses between parents in DC4 do result in low levels of heterosis. But high heterosis in crosses between parents from DC3 and DC2 are quite frequent. This result helped further in attaching weights [DC2 = -1, DC3 = DC2 = 1, DC4 = 0] for the various divergence classes from the sole point of view of heterosis.

A procedure for giving alignment scores was evolved. Incorporating the weights for divergence classes, genetic potential scores were computed. Using their mean and standard deviation, four genetic classes were formed in descending order of importance [for details, refer Arunachalam *et al.,* 1998].

4.3.2 Molecular divergence grouping :

Nei's distance and coefficients of similarity provided an analogue to measuring the divergence by D^2 statistic. Using the method developed in Sokal and Sneath (1963), a hierarchical diagram called 'dendrogram' is constructed on similarity indices; of the many methods available , the unweighted pair group method on arithmetic averages (UPGMA) for which softwares are available, is most popular.

For example, genetic divergence between 18 scented rice germplasm was studied using RAPD markers (Raghunathachari *et al.,* 2000). Jaccard's similarly coefficients for RAPD bands between each pair of accessions were used to construct a dendrogram using UPGMA through NTSYS - PC computer package. The method consists in examining pair wise distances and connecting successively the pairs on a norm to generate a tree diagram [for details, refer Sokal and Sneath, 1963]. Though this diagram illustrates the divergence between varieties and between groups, a clear picture of the number of distinct groups, the varieties contained in them and the degree of divergence between them are not readily visible. While sketchy inferences are possible, a grouping analogous to genetic classes using D^2 appears to be a hard access. In the published study, the method enabled an evolutionary analysis of genetic divergence betwen the 18 rice accessions. But direct leads to breeding for trait improvement are, at best, latent in the paper. Regardless, the relevance of marker-assisted breeding for QT improvement is an open question, as discussed in detail elsewhere [see Arunachalam and Chandrashekaran, 1993].

In contrast, an important analogy could be drawn between QTs and isozyms. 'QTs are the outcome of genes expressed in an environment. The DNA sequence comprising the gene is transcribed into mRNA, the mRNA translated to protein, and the protein assumes proper configuration and localization to exert its effect. The effect exerted by the protein is often due to its enzymatic activity or ability to bind to some cell component. All these stages can, and often are, affected by the environment.' (see Kochert). If, therefore, isozyme pattern can be quantified, an analysis of its variation can reflect the variation in a QT closely associated with it. The prime difference between DNA marker and isozyme marker is that the former is independent of environment while the latter covaries with environment. Therefore, while an association between DNA markers and QT loci is far-fetched, it is logical to expect an association between isozyme QTs and morpho-QTs. One major contribution in this regard is the quantification of isozyme pattern by five QTs - number of bands, relative mobility (RM), relative absorption (RA), the within-lane-deviation (WLD) of RM and WLD of RA (Arunachalam *et al.*, 1996).

Encouraging experimental results were obtained. In an experiment with 31 accessions of *Brassica* with four species and a spectrum of varieties (and also germplasm accessions). A few enzymes like esterase (ES), GOT and Peroxidase - Anodal (PA) and cathodal (PB) were used to evaluate the isozyme marker efficiency in correctly identifying the morphometric grouping based on genetic divergence. When considered along with the diagnostic morpho-traits, the isozyme QTs contributed to genetic divergence up to 30-40%. In particular, the contribution of the two new parameters devised by us, namely, Within Lane Deviation of relative mobility and of relative absorption, was substantial. Incidentally the use of the trait, relative absorption, was also novel in the sense it has rarely been used in earlier literature. The isozyme traits mentioned ranked among the top 5 traits contributing to genetic divergence. The efficiency index of the enzyme combination, ES - PA was quite high (73%) in the experiment conducted in 1995. Even esterase alone had an efficiency of 32% in 1995 and 26% in 1996 [more details in Arunachalam *et al.*, 1998]. The studies emphasize the following :

* The experimental logic, the validity of basic conceptualization and the method of quantitative evaluation are more crucial than emphasis on numerous markers alone.

* Isozymes are not highly favoured as markers on the pretext that they are far fewer in number than molecular markers. The experiment with *Brassica* showed that this argument was misplaced.

* The fact that isozyme markers correspond only to coding regions of DNA is a strength and not a weakness, as the example showed. Their direct association with expressed morpho-QTs resulted in their high efficiency in predicting QT variation.

But the analysis of isozyme and morphological QT data involves complex multivariate methods. A software has therefore been developed; starting from raw

data, it forms genetic classes, computes efficiency index of markers in predicting morpho-QT variation, and locates varieties that are correctly identified by markers in each genetic class. Such a diagnostic analysis with a particular crop in a particular envifronment will help to select isozymes for marker-aided plant improvement programs in that crop.

Summing up, we can observe the following :

* The association between molecular markers (MM) and QTs is difficult to conceptualize, as MMs are independent of environment and cannot covary with QTs. In contrast, isozyme QTs provided an alternative with feasible association with morpho-QTs. Methods have also been developed to identify those, which have high correlation with QTs defining yield performance. In this context, the deficiencies of molecular markers need to be recognized when attempting to use them in breeding methodologies.

* Isozyme variation defined by 5 QTs, as explained earlier, was amenable to all quantitative analysis as was possible with morphological QT variation. Such applications provided leads to breeding processes. One example was the selection of parents on the basis of genetic divergence using isozyme QTs. Crosses between such selected parents were found to be more often heterotic than crosses made between parents selected at random [Aruna Kumari, 1998; Mithra, 1999].

4.4 Concept-driven breeding methods : An example :

More than the charismatic influence of quantitative genetics in the broad area of plant breeding, can anything tangible be provided for practical plant breeders? This has remained a popular and nagging query. Consistent research has provided leads to many reliable approaches and we restrict to the most popular 'pedigree' or 'pure line' breeding. While varietal performance was better defined on multiple QTs, it was not easy to select the best traits for differentiation from the many QTs that have relevance. One way to identify a diagnostic set of traits is by using stepwise regression analysis [see, for further details, Draper and Smith, 1981].

Based on the principles detailed above and years' of research towards efficient breeding approaches, an outline of a breeding procedure is given below :

* Starting from a set of prospective varieties for initiating a pure line breeding program, parents likely to generate heterotic crosses are isolated from divergence classes, DC3 and DC2. Earlier experimental work in various crops has shown that, if a choice of an initiating cross has to be made, the first preference should go to heterotic crosses. Further, growing the progeny of every F_1, heterotic or non-heterotic, would restrict the size of the F_2 population due to constriants on cost, time and experimental area. It would therefore be profitable to effect an early generation selection of heterotic F_1 and raise large F_2s of those few crosses. Earlier work has

shown that the chances of recovering selections in F_3 are much higher in F_2s of heterotic crosses than in others. [Koteswara Rao, 1992].

* Individual plants of the F_2 population are assayed for the diagnostic traits. A selection index is constructed with grain yield [or any other trait targeted for improvement] as the dependent variable, Y and the diagnostic traits as independent variables (X). A 'best' regression line is fitted to provide the selection index. Selection index for each F_2 plant was obtained (as the expected Y value by substituting trait values of X in the regression equation).

The selection index values are arranged in descending order of magnitude to provide a F_2 ranked distribution (FRD).

* The FRD is partitioned into four equal strata, T1, T2, T3 and T4, the top one being T1. Earlier experimental studies have shown that, most often, higher number of selections are recovered in F_3 from the progeny of the top strata T1 or, if inadequate, from T1 + T2. Seldom would there be a need to screen the progeny of T3 and T4.

* All F_2 plants in the strata T1 and T2 are therefore advanced to F_3 progeny rows.

From then on, classical procedures are followed to select desired progeny rows [e.g. family selection and combined selection procedures as in Falconer and Mackey, 1996]. The selected material is advanced to further generations till superior pure lines are recovered. For further reference, we denote this method as Targeted Breeding (TGT).

The gist of the above procedure is the three basic steps : evaluation of parental material, choice of parents for F_1, heterotic F_1 selection, F_2 evaluation and selection of top 25 (or 50)% of ranked F_2 distribution. The quantitative genetics principles behind them suggest that, in the process, potential transgressants/recombinants would be trapped in the F_2. When they are advanced, a high frequency of desirable selections would be recovered in F_3 and further generations.

On the other hand, conventional breeding technologies also outline a procedural need to identify source genetic material, to select parents, to make an initiating cross, to raise a large F_2 population, to select a single plant or cluster of plants, to raise from them F_3 progeny population and continue selection for yield or the desired traits in further generations until a new pure line is homogenized. In those steps, a definitive method of selection at various stages is not inherent and it is left to the discretion of the breeder. We further refer to this method as Tentative Breeding (TNT). The breeding process, TGT differs essentially in that aspect; as explained earlier. A logical estimate of the resulting progress shows it is higher in TGT than in TNT (Fig. 3). The sample illustration in Fig. 3 indicates the following :

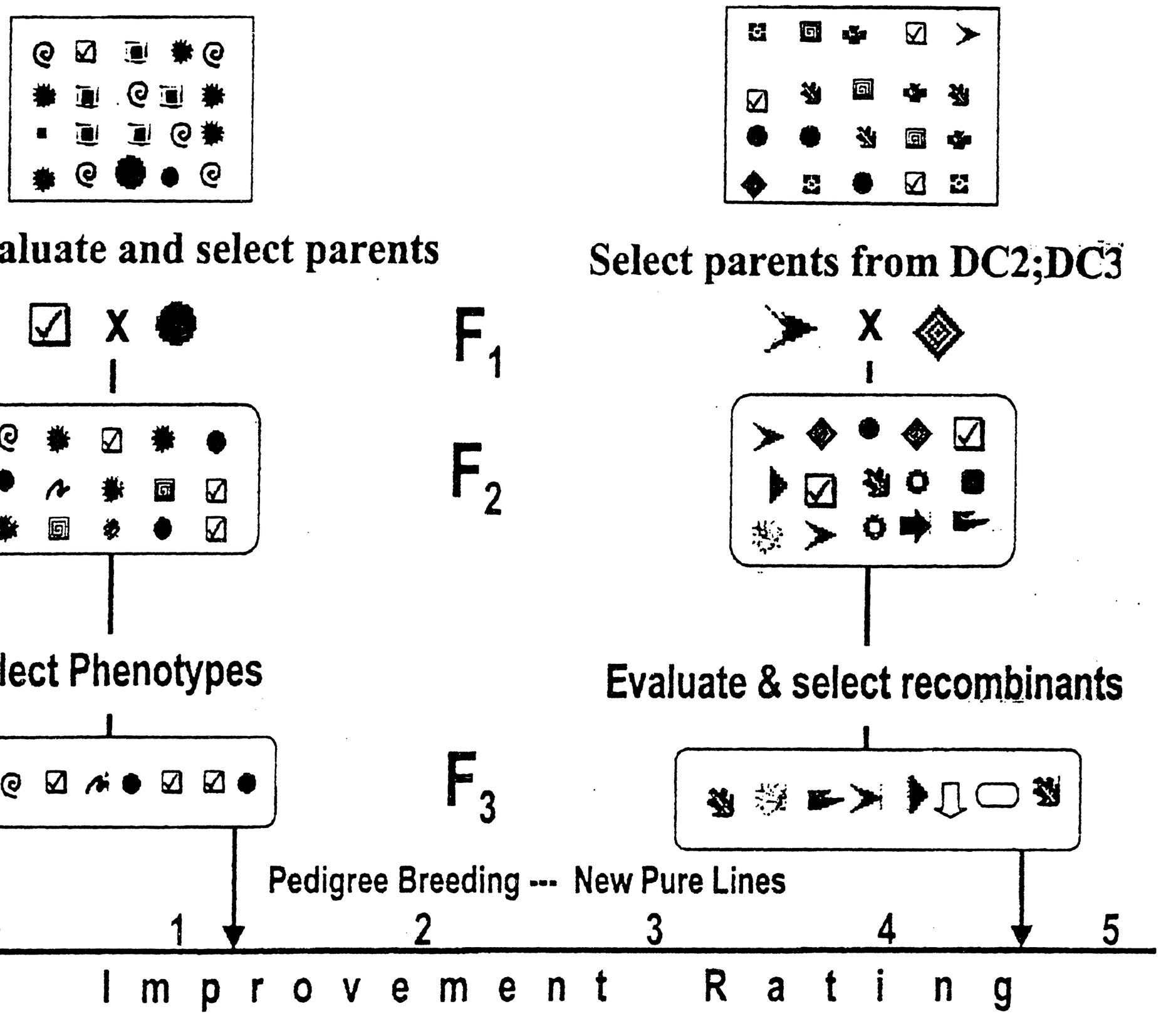


Fig. 3. Expected improvement using tentative and targeted breeding methodology

* The higher frequency of segregants in F_2 indicates that the parental choice in TGT was more appropriate.

* The evaluation and selection process in F_2 have helped TGT to select more number of recombinants than TNT [Note the higher number of recombinants selected by TGT compared to TNT in Fig. 3].

* Further breeding therefore resulted in higher genetic progress in TGT than TNT.

4.5 Mismatch between theory and practice :

It would also be instructive to note how breeding programs do, often, fail to draw visible gains despite recourse to quantitative genetics-assisted initiatives.

Most often quantitative genetics models are applied as cosmetics to a breeding program already formulated on some basis. At times, drifting away from the objective (sometimes not clearly defined) a process is adopted for the sake of applying a

model with articulated results. Many examples could be cited from Indian journals, prominent among them being-raising six generations, P_1,P_2, F_1, F_2, BC_1 and BC_2 to get single gene based parameters or to do joint scaling test etc., a diallel cross analyzed employing a single gene based Vr- Wr graph and also a very general model of Griffing (1956) with a consequent conflict on the choice of parents, multi-point crosses just to estimate some quantitative parameters, forwarding diallel F_1s to F_2 and higher generations and repeating the F_1 diallel analysis on the higher generation data [see also, Arunachalam, 1976].

Breeding is a complex process. The phenotype for QT is governed by a large number of minor genes with small effects but significantly large interactions. Precisely therefore, quantitative genetics uses sophisticated statistical diagnosis of such interactions using defined mating systems and appropriate follow-up in further generaions. Naturally the results of the diagnosis cannot be decisive; being stochastic in nature, and highly modified by environment, QT variation can only be evaluated by specific designs to identify a desirable avenue to advance the material to the next generation. It is therefore mandatory to evaluate the rate, direction and magnitude of progress obtained in the next generation and incorporate appropriate modifications before forwarding them to further generations. Breeding is thus an iterative process, by judicious application of genetic models and appropriate analysis of variation, the process can be strengthened and made more efficient. A well-planned integration of principles and practice can reduce the number of iterations and therefore the breeding time. Essentially this is the gist of a conventional and evolutionary breeding process.

However, molecular breeding whose principal component is marker-aided selection for QT improvement is a new term and a field with supposed strengths over uncertainities built into Mendelian genetics-driven breeding. Though there is the tendency to call the latter as conventional breeding, I perefer it to be termed as classical breeding. The strength of molecular breeding is projected in the following lines :

* Molecular markers (MM) are uninfluenced by environment. They have stable association with loci governing QTs. MMs can be positioned in a conventional map mapping QTs. This will facilitate selection of MMs closely linked to QTLs.

* Because of environmental independence, MMs can be used to select for desired QTLs (and QTs) any time particularly in environments where QTs do not express, e.g. biotic and abiotic stress-related traits.

* For the same reason, marker-assisted selection can bypass effects of genotype X environment interaction and accelerate genetic progress.

* Since MMs are codominant and independent, they are epistasis-free. They are therefore superior in selection for QTs.

4.6 The infirmities inherent in marker-assisted QT improvement :

We have shown that the projected strengths of MM, are their weaknesses in real terms. A marker independent of environment and lacking in expression cannot covary with a QT that is environment-dependent and has an environment-sensitive phenotype. The association of MMs with QTs is, therefore, far-fetched. Crow (2000) in his review on the use of molecular markers for heterosis with QTLs, states that it has not been possible to associate yield changes with specific QTLs which, however, can be achieved by possibly identifying molecular markers associated with each QTL. But it is now known a *priori* how many QTLs constitute a complex trait as yield. Further the number of markers to identify those QTLs would also be quite high and makes the process of marker-assisted selection cost-intensive and arduous. Therefore, a realistic proposition is to establish associations of specific QTLs with specific protein (solute, enzyme) pool and then ultimately associate this metabolic quantity with grain yield expression. The isozyme based quantitative trait exploration mentioned earlier is one step close to that proposition.

In a recent international meet of geneticists, there has been a specific emphasis on the ubiquity and importance of genotype X genotype (epistasis) and genotype X environment interactions in determining patterns of observed phenotypic variation. It was also realized that many QTLs (which are typically large enough when mapped initially to contain anywhere from 400 and 2000 genes) eventually get mapped down to either non-coding upstream regions or intronic region, rather than coding regions of the genome (Joshi, 2002). Therefore the properties of independence from G X E interaction and epistasis do not weight in favour of MMs aiding QT improvement.

The view endorsed by Freeland *et al.*(1999) that 'modern organisms process genotype into phenotype through two distinct stages. First DNA genes are transcribed into RNA messages (mRNA); these messages are then translated into proteins', which subsequently give rise to expression as QTs. In the process of transcription and translation, environment plays a role. Therefore, isozyme variation, which is the expression of protein products of the gene/genes, that is quantified efficiently, can be used as a precocious precusor of the QT variation. It is then logical to postulate an association between isozyme and QT variation. The work on *Brassica* with isozyme QTs mentioned earlier, could distinguish between and within 4 species, select desirable parents, establish heterosis in crosses and recombinants in F_2 (Shefali, 2000; Aruna Kumari, 1998, Mithra, 1999).

How then can we take advantage of modern development of molecular markers? We recall the model, P = G + E. In the model the projected strength of MMs is the characterization of G. Intensive and cohesive research is needed to characterize conceptualization [particularly, $X_{TT} = 1$, $X_{Tt} = 0$,$X_{tt} = -1$] should be abandoned in favour of directed research to conceptualize G on a model of many interacting genes (more than QTLs). Then this concept should be extended to define G on multi traits. Obviously a huge lot of innovative ground remains to be covered.

As subtly expressed by Joshi (2002), molecular genetics has taken a bottom-up approach that has occasionally but relatively rarely climbed up even to the level of whole organism. Quantitative Genetics. has taken to top-down approach,which has rarely managed to go below the level of individual organism. We need a cohesive integration of both the approaches to solve the fundamental and elementary model, P = G + E more efficiently and build on it to deal with realistic and more complex models of P.

4.7 How is Quantitative Genetics positioned today ?

But then, are we in a position to achieve this? Where do quantitative Genetics and those who profess it stand in this context? I re-emphasize what we have stated in a recent paper (Arunachalam *et al.*, 2001):

With the sprinting developments in digital revolution, horizons, though unreachable once, are in our choice to reach. At the high end of the scale we have a modern communication avenue of "virtual instrumentation". At the other end of the scale, we have methodologies for QT improvement attempted in a routine manner. The design of mating and methods of selection that hold the key to targeted QT improvement are directed by experiments done and published elsewhere and limited to the availability of software for analysis of data. Priorities are thus set not by the crucial problems in a crop or a site or an environment but by availability of means to evaluate results and arrive at inferences and applicable protocols.

Thus the big hurdle in the path of innovative and problem-specific research is the highly limited knowledge software of researchers. Most often it is limited to matters of biology and hardly, if ever, there are inclinations to delve into, at least, concerned knowledge areas of other fields (say methods of linkage estimation, knowledge and operational skill of existing computer softwares, knowledge of inter-faces between computer and laboratory instruments etc.). Motivating incentives or compulsions are essential to extend the horizon of already-gathered knowledge. bright students for innovative and future-oriented biological rsearch are the need of the day. Right environment must be created to encourage existing bright scientists to do active and innovative research. Inter-disciplinary interaction would mould incoming young minds towards frontier technologies including needed skills in allied fields. Path-breaking and bold policy decisions are urgent to integrate agriculture with other fields of relevance. Until such time, the minimum we could do is to compulsorily introduce undergraduate and graduate students of biology and agriculture to such inter-disciplinary education and skill orientation. This would empower them to handle modern biology and its developments in a modern way.

It is acknowledged that environment-genotype symbiosis is the key to sustainable livelihood. This fact is equally valid to our Science. Moulding environment would be an avenue to moulding genotypes. A consonant approach towards this goal should be the right path to a rapid progress.

REFERENCES

Arunachalam, V. 1976. Evaluation of diallel crosses by graphical and combining ability methods. *Indian J Genet* **36**: 358-66.

Arunachalam, V. 1977 Heterosis for characters governed by two genes. *J Genet* **63**: 15-24.

Arunachalam, V. 1981. Genetic distance in plant breeding. *Indian J Genet* **41**: 226-36.

Arunachalam, V. 1988. Extending the definition of additive genetic variance to more than one gene - a viewpoint. *J Genet* **67**: 53-61.

Arunachalam, V. and Bandyopadhyay, A. 1984. Limits to genetic divergence for occurrence of heterosis - Experimental evidence from crop plants. *Indian J Genet* **44**: 548-54.

Arunachalam, V. and Chandrashekharan, S. 1993. RFLP approach to breeding for quantitative traits in plants - a critique. *J Genet* **72**: 73-88.

Arunachalam, V. and Owen, A.R.G. 1971. Polymorphisms with linked loci. Chapman & Hall, London.

Arunachalam, V. , Koteswara Rao, M. V. and Bandyopadhyay, A. 1992. A selection strategy securing yield gains in groundnut. *SABRAO J* **24**: 123-31.

Arunachalam, V. , Prabhu K V and Sujata V 1996. Isozyme parameters in genetic evaluation in Brassica - A tribute to Haldane's vision on the synergy between biochemistry and genetics. *Curr Sci* 70: 1062-65.

Arunachalam, V., Prabhu, K. V. and Sujata, V. 1998. Multivariate methods of quantitative evaluation for crop improvement : conventional and molecular approaches revisited. pp. 793-807. In Crop Productivity and Sustainability - Shaping the Future. Proc. 2nd International Crop Science Congress. *Edited by* V L Chopra, R B Singh and Anupam Varma, Oxford & IBH Publishing Co., New Delhi, India, 1111pp.

Arunachalam, V., Prabhu, K. V. and Sujata, V. 2001. Quantitative Trait Improvement : Mendelian and Molecular Approaches. Proc. of the Symp. Hundred years of post-Mendelian Genetics and Plant Breeding - Retrospect and Prospects, November 6-9, 2001, New Delhi.

Aruna Kumari, 1998. Marker-aided selection for quantitative trait improvement in *Brassicas* Unpublished Ph.D. Thesis, Indian Agricultural Research Institute, New Delhi.

Bennett J. H. (eds.) 1965. Experiments in plant hybridization. Oliver & Boyd, London. 95 pp.

Comstock, R. E. and Robinson, H. F. 1948. The components of genetic variance in population of biparental progenies and their use in estimating the average degree of dominance. *Biometrics* 4: 254-56.

Comstock, R. E. and Robinson, H. F. 1952. Estimation of average dominance of genes. In Gowen, J.W. (ed.), Heterosis. Iowa State College Press, U.S.A. Chap 30, 494-516.

Cress, C. E. 1966. Heterosis of the hybrid related to gene frequency differences between two populations. *Genetics* **53**: 269-74.

Crow, J. F. 2000. The rise and fall of overdominance. *Plant Breeding Reviews* **17**: 225-58.

Draper, N. R. and Smith, H. 1981. Applied Regression Analysis. 2nd ed. John Wiley & Sons, New York, U. S. A.

Durga Prasad, M.M.K., Arunachalam, V. and Bandyopadhyay, A. 1985. Diversity pattern elucidating choice of parents for hybridization in varieties of groundnut, *Arachis hypogaea* L. *Trop Agric* **62**: 237-42.

Falconer, D. S. and Mackay, T.F.C. 1996. Introduction to Quantitative Genetics. 4th ed. Longman, UK.

Fisher, R. A. 1941. Average excess and average effect of a gene substitution. Ann. Eugen. (Lond.) **11**: 53-63.

Freeland, S.J., Knight, R.D. and Landweber, L.F. 1999. Do proteins predate DNA? *Science* **286**: 690-92.

Griffing, B. 1956. Concept of general and specific combining ability in relation to diallel crossing systems. *Aust J Biol Sci* **9:** 463-93.

Hayman, B.I. 1954. The analysis of variance of diallel tables. *Biometrics* **10**: 235-44.

Jansen, R. C. and Stam, P. 1994. High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136**: 1447-55.

Jensen, N. F. 1970. A diallel selective mating system for cereal breeding. *Crop Sci.* **10**: 629-35.

Jinks, J. L. and Hayman, B. I. 1953. Analysis of diallel crosses. *Maize Genetics Coop Newsletter* **27**: 48-54.

Jinks, J. L. 1954. The analysis of continuous variation in a diallel cross of varieties. *Genetics* **39**: 767-88.

Joshi, A. 2002. Genomics and evolution. *Curr Sci.* **82**: 122-24.

Kempthorne, O. 1957. An Introduction to Genetical Statistics. John Wiley & Sons, New York.

Kochert, G. (Unpublished). Introduction to RFLP mapping and plant breeding applications. Booklet by the Rockefeller Foundation.

Langham, D. G. 1961. The high-low method of crop improvement. *Crop Sci* **1**: 376-78

Li, C.C. 1955. Population Genetics. The University of Chicago Press, Chicago, U.S.A.

Martin, B., Nienhaus, J., King, G. and Schaffer, A. 1989. Restriction fragment length polymorphisms associated with water use efficiency in tomato. *Science* **243**: 1725-28.

Mather, K. and Jinks, J. L. 1971. Biometrical Genetics. 1st ed. Chapman & Hall, London.

Mather, K. and Jinks, J. L. 1977. Introduction to Biometrial Genetics. Chapman & Hall, London.

Mithra, A. C. R. 1999. Evaluation of yield performance of early generation selections in *Brassica.* Unpublished M.Sc. Thesis, Indian Agricultural Research Institute, New Delhi.

Murty, B.R., Mathur, J.B.L. and Arunachalam, V. 1965. Self-incompatibility and genetic divergence in *Brassica campestris* var. brown *sarson. Sankhya B.* **27:** 271-78.

Murty, B. R. and Arunachalam, V. 1966. The nature of divergence in relation to breeding systems in some crop plants. *Indian J Genet* 26A: 188-98.

Murty, B. R. and Arunachalam, V. 1967. Computer programmes for some problems in biometrical Genetics. I. Use of Mahalanobis' D-square in classificatory problems. *Indian J Genet* **27**: 60-69.

Nei, M. 1972. Genetic distance between populations. *Am Nat* 106; 283-91.

Nei, M. 1987. Molecular Evolutionary Genetics. Columbia Univ. Press, New York, pp. 106-07.

Nei, M. and Li, W.H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci.* **76**: 5269-73.

Raghunathachari, P., Khanna, V. K., Singh, U. S. and Singh, N. K. 2000. RAPD analysis of genetic variability in Indian scented rice germplasm (*Oryza sativa* L.). *Curr Sci* **79**: 994-98.

Rao, C. R. 1952. Advanced Statistical Methods in Biometric Research. John Wiley & Sons, New York.

Shefali Bharti. 2000. Quantitative trait fingerprinting of genetic accessions in *Brassica,* Unpublished Ph.D. Thesis, Agra University, Agra.

Singh, R. K. and Chonduary, B. D. 1983. Biometrical methods in Quantitative Genetic Analysis. Kalyani Publishers, New Delhi.

Sokal, R. R. and Sneath, P. H. A. 1963. Numerical Taxonomy, W. H. Freeman & Co., San Francisco, U.S.A.

Vairavan, S. Siddiq, E. A., Arunachalam, V. and Swaminathan, M.S. 1973. A study of the nature of divergence in rice from Assam and Northeast Himalayas. *Theor Appl Genet* **43**: 213-21.

Wang, Z.Y., Second, G. and Tanksley, S.D. 1992. Polymorphism and phylogenetic relationships among species in the genus *Oryza* as determined by analysis of nuclear RFLPs. *Theor Appl Genet* **83**: 565-81.

Zhang, Q., Gao, Y.J., Yang, S.H., Ragab, R. A., Maroof, M.A.S. and Li, Z.B. 1994. A diallel analysis of heterosis in elite hybrid rice based on RFLPs and microsatellites. *Theor Appl Genet* **89**: 185-92.