Theor Appl Genet (1986) 72:377–383

A mixed model for the effects of single gene, polygenes and their interaction on quantitative traits

1. The model and experimental design *

Y. Elkind and A. Cahaner

The Hebrew University of Jerusalem, Faculty of Agriculture, P.O. Box 12, Rehovot 76-100, Israel

Received September 30, 1985; Accepted December 28, 1985
Communicated by A. Robertson

Summary. A model for the effects of single gene (SG), polygenes (PG) and their interaction on quantitative traits was developed. It is a mixed model where the SG is a fixed effect and the PG is a random effect. A two-way factorial experiment, in which the SG and the PG are the main effects, is proposed. The experimental material is comprised of F_3 families derived from F_2 plants heterozygous for the SG. For this experiment an ANOVA table with expected mean square is proposed, which facilitates estimation of the components of the model and testing of their significance. A detailed method for the interpretation of results from such an experiment is proposed, with emphasis on the analysis of the SG×PG interaction. Theoretical and applied aspects of SG×PG interaction is discussed.

Key words: Single gene – Polygenes – Genetic interaction – Mixed model – Experimental design – Quantitative trait

Introduction

Many quantitative traits are known to be controlled by one or more major genes as well as by polygenes (PG) (Falconer 1981; Mather and Jinks 1971), the former having a relatively large effect on the trait under study and the latter, a small effect (Thoday 1977). In this paper the term 'single gene' (SG) is used, and refers to any identifiable gene, irrespective of its relative effect. Investigations concerned mainly with major genes often refer to polygenes as 'genetic background' or 'modifier genes' (Falconer 1981; Mather and Jinks 1971).

Several experimental methods for estimating the effects of SG on quantitative traits have been reviewed by Gilbert (1985 b), some of which also estimate the PG variation. None are capable of estimating the SG×PG interaction, as they assume this component to be non-existent, but evidence for such interaction has, however, been reported (Habgood and Uddin 1984; King 1955; Vasal et al. 1979).

Many SG are of great importance in breeding programs as they have a desired effect on important characters, e.g. *'opaque-2'* which improves the nutritional value of maize kernels (Vasal et al. 1979), *'nor'* which increases the shelf-life of the tomato fruit (Kopeliovich et al. 1979; Ng and Tigchelaar 1977) and *'Rht1'* and *'Rht2'* which reduce the height of wheat plants (Gale 1979). However, such SG frequently have undesirable effects on other important traits which are also controlled by PG. In many cases these effects threaten to jeopardize the SG utilization. Vasel et al. (1979) demonstrated that it is possible to overcome the adverse effect of *'opaque-2'* on grain yield by selection for favourable modifier genes. Similarly, it was suggested that the adverse effect of *'lys3'* in barley could be overcome by manipulation of the genetic background (Tallberg 1982). This procedure may help in overcoming the adverse effects of *'nor'* on fruit color and taste in tomato, and facilitate its utilization in breeding for fruit firmness (Kopeliovich et al. 1979; Ng and Tigchelaar 1977). Thus, information concerning the interrelation between the SG and the PG is important in many cases for breeding.

The purpose of this paper is to propose a model which combines the effects of a single gene as a 'fixed effect', polygenes as a 'random effect', and elucidates the interaction between them. An experimental design which facilitates the estimation of components of the model is suggested, and the methods for interpreting the results of such an experiment are given. Special emphasis is put upon detailed analysis of the interaction.

The genetic model

The genetic effects on the trait under study are the SG, PG and the SG×PG interaction.

* This paper is part of a Ph.D. Thesis of the senior author to be submitted to the Hebrew University of Jerusalem

The single gene

The SG is assumed to have two alleles (*S, s*) and three genotypes (*SS, Ss, ss*). No assumptions are made regarding the degree of dominance. The effect of the SG can be estimated from the means of the groups of individuals sharing the same SG genotype. The individuals' single gene genotype (SGG) should be clearly identifiable by its effect on the character being studied, or on other traits, through pleiotropy or close linkage to one or two marker genes. In addition, progeny tests may be used to distinguish between the heterozygotes and the dominant homozygotes. The effect of the SG is 'fixed' as our aim is to estimate the means of the three possible 'levels' (genotypes) of the SG (Searl 1971).

The polygenes

The PG in this model are numerous enough to warrant the continuous and normal distribution of the genotypes in the population. No assumptions are made as to dominance or other interactions between PG. The variation among the PG is obtained by random sampling of groups of relatives or families from a population segregating for the PG. Thus, the families are a random sample of all possible combinations of the PG, and the family mean represents a certain polygenic constitution of its common ancestor. The PG in the model is considered a 'random effect', as our aim is to estimate the variance of the PG combinations under study.

The interaction between the single gene and the polygenes

In cases where the effect of the SG differs in various PG constitutions, as expressed in different families, and similarly, where the effect of the PG varies with the SGG, interaction between the SG and some or all of the PG exists. The interaction effect can be studied by a two-way factorial design of the SGG and the families. It is essential that each family should include all levels of the SGG.

The genetic design and statistical procedures

The estimation of the components of the genetic model mentioned above is achieved by using a two-way factorial design. One factor is the PG variation, obtained from differences between families, and the other factor, orthogonal to it, is the SG with its three possible genotypes. Consequently, the design allows us to identify and quantify the SG×PG interaction.

The individuals in the experiments may either be arranged at random, or in two levels resembling a split-plot design. In the latter, the families are arranged in several main plots (replicates), whereas the SGG are arranged in a pattern resembling the secondary plots. To avoid complications in the statistical analysis, a balanced structure is highly recommended. In populations segregating for the SG, the above can be achieved by planting an extra number of individuals in each family, and sampling at random an equal number of individuals for each SGG within each family. A detailed example of an experiment suitable for self-pollinating plants, which follows the above structure, is given below.

F_3 families derived from single gene heterozygotes

In this experiment the population to be investigated is comprised of individuals, each belonging to a specific F_3 family, which segregates for the SG. Each F_3 family is derived from a single selfed F_2 plant heterozygous (*Ss*) for the SG. The parents of the F_1, from which the F_2 source population arises, must differ in SGG. Furthermore, to obtain maximum divergence in the PG they should differ widely in the PG for the character under study. When the parents are not homozygous in all PG, it is advisable to derive the source F_2 population from a single F_1 individual. The traits under study in individuals of the F_3 generation are measured and their SGG is determined.

When a more detailed investigation of the polygenic system is desired, the parents and F_1 populations of the F_3 under study can be included in the experiment, and each of the above populations should be arranged at random or in split-plots according to the design chosen.

The model of the design and the analysis of variance are presented for both random and split-plot designs.

The random design

Y_{ijk} is the performance of the k-th individual from the j-th family with the i-th SGG.

$$Y_{ijk}=U+S_i+P_j+S\times P_{ij}+W_{ijk}, \quad (1)$$

where U is the general mean, S_i is the SG effect of the i-th genotype, with three possible genotypes, (i=1 (*SS*), i=2 (*Ss*), i=3 (*ss*)), P_j is the effect of the j-th family (j=1 to p, the number of families sampled), $S\times P_{ij}$ is the effect of the interaction between the j-th family and i-th SGG, W_{ijk} is the effect of the k-th individual within the ij-th SG×PG combination (k=1 to n). S_i is a 'fixed effect' with a zero mean; P_j is a random effect with a zero mean and variance σ^2_p; S×P has zero mean variance σ^2_{sp} and for each family the sum of the SG effects is zero (Scheffe 1959).

The split-plot design

Y_{ijlk} is the performance of the k-th individual from the l-th plot of the j-th family with the i-th SGG.

$$Y_{ijlk}=U+S_i+P_j+R_{jl}+S\times P_{ij}+S\times R_{ijl}+W_{ijkl}. \quad (2)$$

Where U, S_i, P_j, $S\times P_{ij}$, are as in (1) with p families, each having r plots with n individuals for each SGG within each plot. R_{jl} is the random effect of the plots within each family having a zero mean and σ^2_r variation. $S\times R_{ijl}$ is the interaction between the i-th SGG and the l-th plot within the j-th family (l=1 to r). It has zero mean, variance σ^2_{sr}, and for each plot within family the sum of the SG effects is zero (Scheffe 1959). W_{ijkl} is the effect of the k-th individual within the j×l-th plot and the i-th SG combination, random in nature with a zero mean and σ^2_w.

Interpretation of results

General

There are a number of possible paths in interpreting the results, depending on the presence and nature of

Table 1. ANOVA for F_3 families in a random design

Source		df	E(MS)[a]
Mean	(U)	1	$\sigma_w^2 + 3n\sigma_p^2 + 3pnU^2$
Between F_3 families	(P)	p−1	$\sigma_w^2 + 3n\sigma_p^2$
Single gene	(S)	2	$\sigma_w^2 + n\sigma_{sp}^2 + (pn/2)\,\Sigma\,Si^2$
Interaction	(S×P)	2 (p−1)	$\sigma_w^2 + n\sigma_{sp}^2$
Within families-SSG	(W)	3p (n−1)	σ_w^2

[a] The expected mean square derived following Cornfield and Tukey (1956)

Table 2. ANOVA for F_3 families in a split-plot design

Source		df	E(MS)[a]
Mean	(U)	1	$\sigma_w^2 + 3n\sigma_r^2 + 3nr\sigma_p^2 + 3pnrU^2$
Between F_3 families	(P)	p−1	$\sigma_w^2 + 3n\sigma_r^2 + 3nr\sigma_p^2$
Plots within families	(R)	p (r−1)	$\sigma_w^2 + 3n\sigma_r^2$
Single gene	(S)	2	$\sigma_w^2 + n\sigma_{sr}^2 + nr\sigma_{sp}^2 + (pnr/2)\,\Sigma\,Si^2$
Interaction	(S×P)	2 (p−1)	$\sigma_w^2 + n\sigma_{sr}^2 + nr\sigma_{sp}^2$
Interaction	(S×R)	2p (r−1)	$\sigma_w^2 + n\sigma_{sr}^2$
Within plots SGG	(W)	3pr (n−1)	σ_w^2

[a] The expected mean square derived following Cornfield and Tukey (1956)

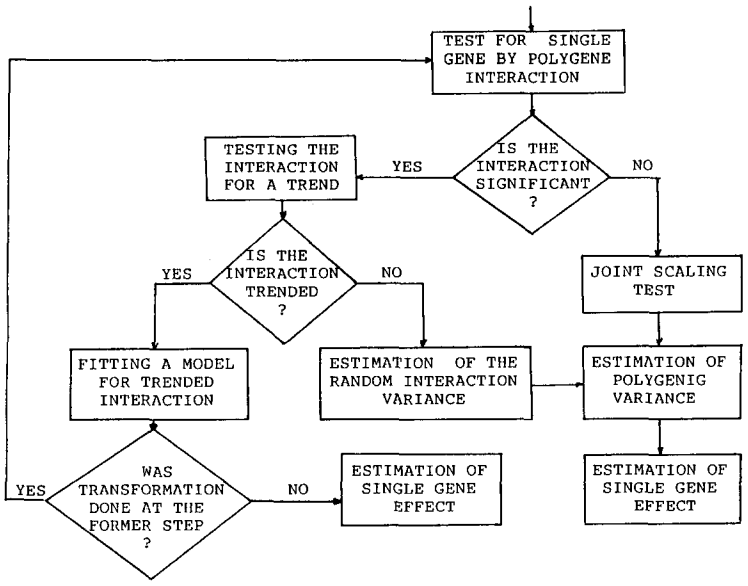

Fig. 1. Flow chart for the interpretation of the experimental results of F_3 families derived from single gene heterozygotes

the SG×PG interaction. These are described in Fig. 1 and the use of graphical representations as well as of statistical analyses are illustrated and their importance discussed.

1 Test for single gene×polygenes interaction

The significance of the SG×PG interaction is tested by the use of ANOVA (Table 1 or 2). First, the assumptions of the ANOVA should be tested, in particular normality and homogeneity of variance (Scheffe 1959; Sokal and Rohlf 1981). In the random design the F test for the interaction is performed on the mean square of the interaction (S×P) divided by the 'within families SG' mean square (W). In the split-plot design, the mean square of the interaction between the SG and 'plots within families' (S×R) gives the appropriate error for the F test of the SG×PG interaction.

2 Interpretation of interaction

Testing the interaction for a trend. The nature of the interaction may be random or follow one of various trends. Analysis of its nature should therefore employ

graphical representations as well as model-fitting techniques. The graphical presentation in Fig. 2B is an example of random interaction. Examples of interactions which follow certain trends can be seen in Figs. 2C, D. It appears that the graphical presentation is the best tool to determine whether the interaction is random or not, although some statistical procedures can be followed, e.g. rank correlation or Runs Tests (Sokal and Rohlf 1981). Theoretically, the interaction can both partially be random, and follow a certain trend (Mather and Jinks 1977). Therefore, if no clear-cut answer regarding a trend in the interaction is apparent, an attempt to fit a model for trended interaction is recommended. Furthermore, it is possible that not all three SGG are involved in the interaction, which can be detected by removing the SGG, one at a time, from the ANOVA. This approach can also be used in the following steps of the analysis of the interaction.

Fitting a model for trended interaction. The objectives of fitting a model for the interaction are to understand the nature of the interaction, as well as to predict its effect. These two objectives, however, may at times yield different models. In the former, a biologically sound model is fitted through which the nature of the interaction can be understood, as for example, the use of a log transformation or a square root (Falconer 1981; Mather and Jinks 1977). The prediction objective may yield a model which will give the best fit, yet have no clear-cut biological interpretation.

When interaction is due to the lack of additivity of the SG and the PG, but the means of the SGG are in consistent order within the families, which are ranked in the same order in all three SGG, transformations can serve as a tool to remove the interaction and also model it (Scheffe 1959). This information can be obtained either from the graphs or from rank correlations between the SGG along the families. The possible transformations should be strictly increasing, in order to preserve the original relative rank of the data (Scheffe 1959). A transformation is chosen which either removes the interaction, or reduces it while leaving the residual interaction random.

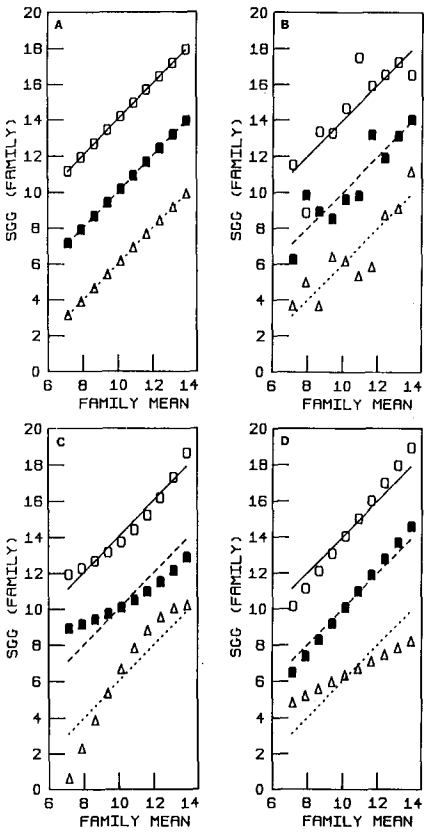

Fig. 2A–D. Expected results for the F_3 families experiment. The SG×PG interaction is absent (**A**) or random (**B**), with a non-linear (**C**) or linear (**D**) trend. The means of the within family SGG, *SS* (□), *Ss* (■), *ss* (△), are plotted against the family mean; the expected means without interaction are the lines (——), (– – –) and (· · · · ·), respectively

Random interaction. When the interaction is random in nature, it is meaningful to estimate its variance, from the ANOVA (Table 1 or 2), where σ^2_{sp} is the estimate of the interaction variance. Its relative magnitude, in comparison with other sources of variance, can express the importance of the interaction in addition to its statistical significance. The variance between and within families in each SGG should then be tested for homogeneity. If not homogeneous, variance components should be estimated separately in each of the three SGG. If there is homogeneity, then σ^2_p and σ^2_w, the between and within families variances, can be used for further polygenic interpretations.

3 Analysis of polygenic variation

The analysis of PG can be performed according to conventional methods (Mather and Jinks 1971; Mather and Jinks 1977; Cahaner and Hillel 1980). Here, only those points which need modification will be presented. The Joint Scaling Test can be performed only if the parents and the F_1 were included in the experiment. Otherwise, estimation of components of genetic variance and heritability is based on the assumption that no interaction between the polygenes exists.

Joint scaling test. For the Joint Scaling Test (Cavalli 1952; Mather and Jinks 1977), modified coefficients should be used to incorporate the single gene into the model. These are shown in Table 3, where d*, h*, D*, H*, are the additive and dominant effects for the SG and for the PG, respectively.

Estimating the components of genetic variance and heritability. The components of genetic variance can be estimated by using the variance between and within the

Table 3. The coefficents for Joint Scaling Test for F_3 population

Single gene	Population	Coefficents				
		m	d*	h*	D*	H*
SS	P_1	1	1	0	1	0
ss	P_2	1	−1	0	−1	0
Ss	F_1	1	0	1	0	1
SS	F_3	1	1	0	0	1/4
Ss	F_3	1	0	1	0	1/4
ss	F_3	1	−1	0	0	1/4

F_3 families which can be obtained from the ANOVA table. Following Mather and Jinks (1977) it can be stated:

Variance between F_3 families:

$$V_{1F_3} = 1/2\ D + 1/16\ H = \sigma^2_p. \quad (3)$$

Variance within F_3 families:

$$V_{2F_3} = 1/4\ D + 1/8\ H + Ew = \sigma^2_w. \quad (4)$$

A heritability estimate (in between 'narrow sense' and 'broad sense') can be obtained following Cahaner and Hillel (1980).

$$h^2_{F_3} = 3\sigma^2_p/2(\sigma^2_w + \sigma^2_p) \quad (5)$$

The environmental variance, Ew, can be estimated from the variance within the parents and the F_1 populations if they were included in the experiment, as seen in (6).

Environmental variance:

$$Ew = 1/4\ V_{P_1} + 1/2\ V_{F_1} + 1/4\ V_{P_2}. \quad (6)$$

A perfect fit solution can yield estimates of the variance components D and H from (3), (4) and (6), facilitating estimates of the 'narrow sense' and 'broad sense' heritability values.

4 Estimating single gene effects

The mean of each SGG is calculated with its standard error $SE = (MSW/pn)^{0.5}$ when there is no interaction, and $SE = (MSSP/pn)^{0.5}$ when there is significant interaction of random nature. In the absence of SG×PG interaction, the probability that an individual of a given SGG will have certain larger or smaller values, can be calculated by using the SGG mean and σ^2_w from the ANOVA. In the presence of random interaction the variance estimate used is σ^2_{sp}.

In cases where the interaction is trended, and a suitable transformation which removes the trend is not found, a generalized interpretation of SG effect should be avoided. Any conclusion as to the effect of the SG should be drawn separately for each family, or groups of families, as is done in the fixed effect model with significant interaction.

In simple cases, when a straightforward interpretation of the transformation is possible, the conclusion can be based on the transformed data, and the chosen transformation may indicate the nature of the interaction. If for example, a log transformation was chosen, the multiplying factor of each genotype can be estimated from the antilog of the SGG mean. In more complicated transformations, however, the transformed variables should be used for the analysis and for the prediction of certain values, such as the confidence interval or selection limits. These values should then be subjected to reverse transformation in order to allow prediction in terms of the original variable units.

Discussion

The mixed model for SG, PG and their interaction was presented, and an experiment, comprised of F_3 families derived from SG heterozygotes, was suggested in order to estimate the components of the model and to test their significance. A detailed interpretation of results was proposed, with special emphasis on analysis of the interaction.

The experimental population

The model can be applied to organisms with mating systems other than self-pollinating plants, e.g. cross-pollinating plants, dioecious animals and plants, and haploid organisms. It can also be applied to populations other than F_3 families, such as half-sibs, full-sibs, F_4 families segregating for the SG, F_2 of North Carolina design II, double cross, etc. (Mather and Jinks 1971). These populations can be used if they provide the two-way factorial layout, which is essential for the evaluation of the interaction.

For maximum reliability and relevance of the estimates of the SG×PG model, the segregating population should contain the maximum relevant PG variation. In the F_3 families experiment this can be achieved by choosing parents with maximum divergence for the initial cross. To ensure that the SG has two segregating alleles, the parents should be homozygous for alternative alleles of the SG, or, if not available, the F_2 population should be derived from a single heterozygote (*Ss*) F_1 plant.

Inclusion of the parents and F_1 populations in the experiment facilitates a more detailed polygenic analysis, which includes the Joint Scaling Test (Cavalli 1952); and estimation of additive and dominance genetic variance components, as well as the environmental variance component (Mather and Jinks 1977).

Their inclusion should depend on the objectives of the research, since they are not essential for the estimation of the interaction and the SG effect.

Organizing the experiment

The random design yields a better estimate of the component of variance between families, whereas the split-plot design has greater ability to detect SG×PG interaction. The split-plot design also provides an estimate for the interaction between the SG and plots within families, which is also an estimate of the SG by environment interaction. Furthermore, a split-plot experiment is technically easier to perform, and in most cases resembles, more closely, the conditions under which plant breeding activities are carried out. Both designs can also be arranged in blocks, which facilitates detection of other interactions, such as families by block, in addition to reduction in the residual variance. Further theoretical and simulation studies are required to explore the relationship between the genetic parameters, the experimental size, and the power of the analysis.

The single-gene polygene interaction

The SG×PG interaction is included in the suggested model, and its presence tested by the proposed experiment. Furthermore, its nature may be analyzed by using the suggested interpretation of results. This component was not included in other models suggested for the estimation of major gene effects and polygenic variation, as its absence was assumed (Gilbert 1985b; Hasstedt 1982; Morton and MacLean 1974). Gilbert (1985a) did not attempt to use available methods (Mather and Jinks 1977) to detect non-allelic interactions.

The current models and experimental design dealing with genetic interaction for quantitative traits (Mather and Jinks 1971) estimate the average effects and variance of epistasis, which is usually separated into three components: the additive by additive, the additive by dominance, and the dominance by dominance. The ability to study the involvement of the SG in the genetic interaction in detail, as presented in this paper, may improve the understanding of genetic interaction of quantitative traits.

Estimating the effect of a single gene

The generality of the estimate of the SG effects largely depends on the variation in the polygenic constitutions on which the estimate is based. A widely used method to obtain those estimates is that of the isogenic-lines, in which there is only one polygenic constitution. However, generalized conclusions are usually drawn on the effect of the SG in the species. Although this is appropriate only in the absence of interaction between the PG and the SG, this condition is seldom checked. The suggested experimental procedure, and the underlying model, facilitate estimation of the SG effect over a wide range of polygenic constitutions, thus giving a more robust estimate. The generality of the SG estimates in the suggested experiment depends on how well the relevant PG variation is represented in the F_2 base population. Other genetic designs, such as double-cross, may produce a wider PG variation, and hence yield a more generalized estimate for the SG.

The suggested experiment provides an appropriate test for the reliability of the SG effect, as it is able to detect and measure the interaction. In the case of random interaction, its mean-square serves as an appropriate error term for the SG effects. In the case of trended interaction, there is no certain reliable SG estimate, as its effect varies in different PG constitutions.

As a method to estimate SG effects, the suggested experimental procedure has also a practical advantage over the isogenic-lines, since the former requires only three generations to reach the experimental phase, while the latter needs at list six generations to establish isogenic-lines (BC_5).

Possible application of the model

Information about the interrelation between the SG and PG obtained from the suggested experiment could assist in choosing the breeding strategy for traits affected by both, and could indicate whether one could overcome or improve the effect of the SG by manipulating the PG. Furthermore, the analysis of the interaction would show in which SGG selection for the desired PG should be performed. In the absence of interaction, the selection would be equally efficient in all three SGG. On the other hand, when interaction is present, optimal selection for the desired PG should be performed in a specific SGG.

Acknowledgements. I wish to thank Prof. N. Kedar for suggesting the biological problem treated and for his valuable advice; to M. Friedmann for useful discussions and assistance; to Prof. A. Ashri for his encouragement and helpful remarks and to Prof. M. Soller for his constructive comments.

References

Cahaner A, Hillel J (1980) Estimating heritability and genetic correlation between traits from generations F_2 and F_3 of self-fertilizing species: a comparison of three methods. Theor Appl Genet 53:33–38

Cavalli LL (1952) An analysis of linkage in quantitative inheritance. In: Reeve ECR, Waddington CH (eds) Quantitative inheritance. HMSO, London, pp 135–144

Cornfield J, Tukey JW (1956) Average values of mean squares in factorials. Ann Math Stat 27:907–949

Falconer DS (1981) Introduction to quantitative genetics. Longman, London

Gale MD (1979) The effects of 'Norin 10' dwarfing genes on yield. In: Ramanujam S (ed) Proc 5th Int Wheat Genet Symp, Vol 2. Indian Soc Genet Plant Breed, New Delhi, pp 978–987

Gilbert DG (1985a) Estimating single gene effects on quantitative traits. 1. A diallel method applied to Est 6 in *D. melanogaster*. Theor Appl Genet 69:625–629

Gilbert DG (1985b) Estimating single gene effects on quantitative traits. 2. Statistical properties of five experimental methods. Theor Appl Genet 69:631–636

Habgood RM, Uddin MR (1984) The effect of genetic background on grain shrivelling in lys 3 barley, and its possible modification by recurrent selection programme. Z Pflanzenzücht 93:65–77

Hasstedt SJ (1982) A mixed-model likelihood approximation on large pedigrees. Computers and biomedical research 15:295–307

Kearsey MJ, Jinks JL (1968) A general method for the detection of additive, dominance and epistatic components of variation. 1. Theory Heredity 23:403–409

King JWB (1955) Observations on the mutant 'pygmy' in the house mouse. J Genet 53:487–497

Kopeliovitch E, Rabinowitch HD, Mizrahi Y, Kedar N (1979) The potential of ripening mutants for extending the storage life of the tomato fruit. Euphytica 28:99–104

Mather K, Jinks JL (1971) Biometrical genetics, 2nd edn. Chapman and Hall, London

Mather K, Jinks JL (1977) Introduction to biometrical genetics. Chapman and Hall, London

Morton NE, MacLean CJ (1974) Analysis of family resemblance. 3. Complex segregation of quantitative traits. Am J Hum Genet 26:489–503

Ng T, Tigchelaar EC (1977) Action of the non ripening (nor) mutant on fruit ripening of tomato. J Am Soc Hortic Sci 102:504–509

Scheffe H (1959) The analysis of variance. Wiley and Sons, New York

Searle SR (1971) Linear models. Wiley and Sons, New York

Sokal RR, Rohlf FJ (1981) Biometry. Freeman, San Francisco

Tallberg A (1982) Characterization of high-lysine barley genotypes. Hereditas 96:229–245

Thoday JM (1977) Effects of specific genes. In: Pollak E, Kempthorne O, Bailey TB (eds) Proc Int Conf Quant Genet. Iowa State Univ, Ames, Iowa, USA, pp 141–159

Vasel SK, Villegas E, Bauer R (1979) Present status of breeding quality protein maize. In: Int Symp Seed Protein Improv Cereals and Grain Legumes. IAE, Vienna, pp 127–150