ON METHODS FOR DETECTING LINKAGE FROM ANALYSIS OF GENETIC COVARIANCE

A thesis

Submitted to the Faculty of Graduate Studies and Research

in Partial Fulfilment of the Requirements

for the Degree of

Doctor of Philosophy

in the

Department of Crop Science and Plant Ecology

University of Saskatchewan

by

RONG-CAI YANG

Saskatoon, Saskatchewan

January 1989

Abstract

One of the critical issues in designing crop improvement programs is whether or not linkage is important in quantitative traits. This thesis is concerned with evaluating methods for detecting and estimating linkage among genes controlling a quantitative trait. Gates' test for linkage was evaluated by theoretical analysis and computer simulation. A general expression for variances and covariances between relatives in selfed generations derived from a cross of two inbred lines was developed for a two-locus model with both linkage and epistasis. From analysis of this general expression, epistasis was found to mimic the effects of repulsion linkage. Furthermore, the use of an approximate expectation for variances and covariances in Gates' test caused an upward bias in both additive and dominance effects involving linkage.

Twenty genetic models were simulated with eight values of both coupling and repulsion linkages to determine under which genetic situations Gates' test is a valid test for linkage. Coefficients of determination (R^2), calculated as the proportion of total variability among 30 variances and covariances in Gates' test explained by additive and dominance effects only, were close to unity for most combinations of models and linkage values. R^2 values in repulsion linkages showed greater departure from unity than those in coupling linkages in both the presence and absence of epistasis. Although Gates' test generally fails to detect linkage because of high R^2 values, it is more sensitive to repulsion linkages than coupling linkages.

Numbers of families in hierarchical mating designs were the single most important factor in determining precision of the estimated variances and covariances used in Gates' test for linkage. This number must be at least several hundred to give suitably precise estimates of variances and covariances for traits with low heritability. Numbers of sub-families and numbers of replications can be of minimum size (about two).

Weighted least squares and maximum likelihood methods were used to fit linkage, epistasis or genotype-environmental interaction models to data from a two-year field evaluation of hierarchical progenies in two spring wheat crosses, Potam × Ingal and RL4137 × Ingal. The genotype-environmental interaction model, which allows for heterogeneous error variances, was most satisfactory. The environmental variation between the two years may be due to the different climatological patterns. Among components of genetic variance, additive variance was of principal importance for all eight traits measured in both Potam × Ingal and RL4137 × Ingal.

Acknowledgements

The author is greatly indebted to Dr. R.J. Baker for suggesting the thesis topic and for providing an intelligent environment which is necessary in the field of research. His guidance throughout the study and his constructive criticism in the preparation of this manuscript are very much appreciated. It has been a distinct privilege to work with Dr. Baker.

Dr. B.G. Rossnagel is thanked for his advice and help with the field experiments when Dr. Baker was on sabbatical leave from the University of Saskatchewan. Thanks are also due to Mr. W. Dougall for his technical assistance.

The author wishes to express special thanks to his wife, Xiao-yan, for her help and for providing considerable support and encouragement when this thesis was being written.

Financial support was provided through a Natural Sciences and Engineering Research Council grant to Dr. R.J. Baker and a University of Saskatchewan College of Graduate Studies and Research scholarship to the author.

Table of Contents

List of Figures

List of Tables

1. Introduction

Quantitative genetic studies have been directed at many agronomic traits of interest to the plant breeder. Because of the relatively large environmental effects as well as the large number of loci involved, data for these traits often appear in continuous or quantitative form. In order to make genetic inferences from quantitative data the plant breeder must make use of statistics, such as the mean, variance and covariance, to describe this variation.

With the presence of a large number of genes, linkage is usually important in interpreting these variances and covariances, especially in populations where genotypic frequencies are not in linkage equilibrium. Knowledge of the type and amount of gene action and the degree of linkage operating in a quantitative trait is important to the plant breeder because it influences the choice of an appropriate breeding procedure for crop improvement. For example, if coupling linkage is predominant, then the plant breeder can maintain this linkage block of favorable genes by strictly selfing methods; if, on the other hand, repulsion linkage is important, then the plant breeder would like to break this linkage block by a few generations of random mating to obtain the favorable gene combination.

In the search for superior genotypes, the plant breeder frequently classifies the material of a generation into groups according to ancestry of

the group. Groups of one generation may be related to groups of another generation through a common ancestor in a preceding generation. Thus, variances and covariances between relatives can be estimated. The expectations of these observed variances and covariances are a linear function of various components of genetic variance and environmental variance. Components associated with gene action and linkage can be estimated by appropriate statistical methods (e.g., the least squares method).

Gates *et al.* (1960) presented a procedure for detecting the presence of linkage among genes controlling a quantitative trait. The method is based on the analysis of a set of variances and covariances among selfed relatives. The major objective of this thesis was to evaluate Gates' test for linkage. Theoretical analysis, based on Hayman's (1954, 1955) device and computer simulation, was employed. The second objective of this thesis was to illustrate, through data collected from field experiments, how different methods can be used to estimate linkage, gene action and environmental effects.

2. Literature Review

2.1. The role of linkage in plant breeding

Plant breeding is concerned with the production of improved varieties of crop plants. The task of the plant breeder is to identify and select superior genotypes. For qualitative traits, the genotypes can be identified with little ambiguity through the usual Mendelian analysis (Mather, 1979). However, many agronomic traits of interest to the plant breeder are quantitative. Their study and evaluation depends on measurement instead of classification (Falconer, 1981). Direct Mendelian analysis of quantitative traits is not possible (Mather, 1979). Statistics, such as mean, variance and covariance, are necessary to describe and identify the components of variation for the traits under study (Mather and Jinks, 1982). The different variance components reflect sources of variation arising from gene action and/or linkage effects. Information regarding the kind and amount of gene action or linkage effect is desirable for designing an appropriate breeding scheme.

Response to artificial truncation selection for quantitative traits has been predicted, using the well-known formula (e.g., Cockerham and Matzinger, 1985):

$$\Delta = iC_{XY}/\sigma_X, \tag{2.1}$$

where Δ is the expected response, i is the standardized selection differential,

C_{XY} is the covariance between the genotypic values of response units, Y, and the phenotypic values of related selection units, X and σ_X is phenotypic standard deviation of the selection units.

In general, C_{XY} is the covariance between relatives and its genetic expectation is derived from the consideration of the pedigree connection between X and Y. While both σ_X^2 and C_{XY} will vary with the type of progeny or family under consideration, C_{XY} is of major importance. Thus, the ratio of responses for any two types of families is given approximately by the ratio of the respective covariances (Cockerham and Matzinger, 1985).

With certain assumptions, C_{XY} can be interpreted as a linear function of variance components arising from gene action or linkage effect and will be of primary interest to the plant breeder. For a single locus additive model with dominance, the genetic content of C_{XY} can be easily found (e.g., Kempthorne, 1957). The situation becomes complicated when consideration shifts from single locus to multilocus models. One of these complications is linkage. For quantitative traits, normally controlled by genes at many loci, linkage must be an important factor and must affect covariances among relatives. Linkage causes a lower level of recombination than is associated with unlinked genes. Thus, linkage causes an overabundance of parental combinations and a corresponding deficiency in recombinations (Allard, 1960). As the number of loci affecting a quantitative trait is increased, the likelihood of linkage is increased (Gates *et al.*, 1960).

In discussing issues in quantitative genetics critical to crop improvement, Baker (1984) outlined the general effects of linkage on genetic variances and covariances of quantitative traits. Coupling linkages increase

genetic variances of traits and cause positive covariances between traits; repulsion linkages reduce genetic variances and cause negative covariances (Baker, 1984). The effects are greater in magnitude as linkage becomes tighter. In terms of components of genetic variances and covariances, Moll *et al.* (1964) pointed out that coupling linkages increase additive genetic variance while repulsion linkages reduce it; dominance variance is increased regardless of linkage phases. Knowledge of linkage phases (coupling or repulsion) will provide the plant breeder with a guideline as to whether intermating is necessary to break linkage blocks. This question is particularly important to breeders who plan to use population improvement as a breeding procedure in self-fertilized crops (Knott, 1987).

In self-fertilized crops, it is very costly or even impossible to make substantial numbers of crosses. The effects of intermating may be negated by genetic drift unless at least 30 crosses among randomly chosen F_2's are made in an intermating generation (Baker, 1968). There is a debate on benefits of random intermating before selection in self-fertilized crops from a theoretical (e.g., Baker, 1968; Pederson, 1974; Bos, 1977; and Silvela and Diez-Barra, 1985) and an experimental (e.g., Miller and Rawlings, 1967; Altman and Busch, 1984; and Frederickson and Kronstad, 1985) point of view. The debate centers around the question of whether or not intermating is superior to selfing following selection.

Under a two-locus model with additive genes which combine multiplicatively, Minvielle (1987) found, from analysis of generation means, that reduction in heterosis in F_2 depends on the rate of recombination between the loci. F_1 superiority is nearly halved in the F_2 if linkage is tight

and is the same in the F_2 if genes at different loci are independently segregating. On the other hand, Moll *et al.* (1964) analysed data obtained from F_2 populations of two single crosses between corn inbreds and found that linkage effects cause an upward bias in estimates of dominance variances and average level of dominance. In either case, linkage effects mimic overdominant effects. This phenomenon has been called *pseudo-overdominance* (Falconer, 1981) or *associative overdominance* (Bulmer, 1980).

Study of the effects of linkage becomes more complicated or even unmanagable when epistasis is considered. Linkage and epistasis represent physical and functional aspects of non-alleles (Kojima and Lewontin, 1970). When an effect of interaction between particular non-alleles is favorable under selection, the chance of survival of these alleles is increased in comparison to other non-alleles at these loci which do not interact favorably. In a study on whether recurrent selection (RS) is superior to a self-fertilized system (SS), Silvela and Diez-Barra (1985) showed a clear superiority of RS over SS when repulsion linkage is present. Such superiority becomes even greater when epistatic interaction enhances the selection advantage of new recombinants.

The study of linkage in quantitative traits requires use of biometrical genetic techniques. A general account on these techniques can be found in standard books, such as those by Kempthorne (1957), Crow and Kimura (1970), Mather and Jinks (1971, 1982), Li (1976), Bulmer (1980) and Falconer (1981). Specific accounts on application of quantitative genetics to plant breeding are made by Hallauer and Miranda (1981) on maize breeding,

Baker (1986) on selection index, Wricke and Weber (1986) on selection theory and Mayo (1987) on breeding theory.

2.2. Covariances of relatives

2.2.1. Genetic model

According to Ewens (1979) and Jameson (1977), the first attempt to study covariances of relatives can be traced back to a series of a century-old papers by Sir Francis Galton who suggested the law of ancestral heredity:

> "the mean character of the offspring can be calculated with more exactness, the more extensive our knowledge of the corresponding characters of ancestry."

Galton proposed that the individual was related to each generation of its ancestry by some constant proportion of that in the previous generation. Galton's theory was picked up and expanded by Karl Pearson and his associates. However, this theory denies the Mendelian inheritance and implies that quantitative traits cannot be explained by Mendelian laws. Historically, there was a bitter controversy between the Mendelians led by Bateson, and the biometricians led by Pearson.

Reconciliation between the two groups was made not only possible but indeed inevitable by Fisher (1918) in his landmark paper entitled "The correlation between relatives on the supposition of Mendelian inheritance". By assuming that genes controlling quantitative traits such as human height are inherited in the Mendelian fashion in a large random mating equilibrium population, Fisher concluded that covariances between various types of relatives such as father and son could be expressed as a linear function of

parameters that are now known as additive variance (σ^2_A) and dominance variance (σ^2_D). He also gave a general consideration of non-allelic interaction (epistasis), but did not incorporate epistasis into formulae for covariances of relatives.

Subsequently, Wright (1921, 1935) considered a "small" non-equilibrium population and proposed a model for covariances in inbred populations. For a single locus, σ^2_A and σ^2_D were redefined for inbred populations since additional terms arise from inbreeding. Wright (1935) also analyzed genetic covariances (or correlation) between relatives for a special epistatic model where the genotypic value is greatest when the number of plus genes is optimized.

Malecot (1948, 1969) considered inbreeding in terms of the probability of genes being identical by descent. Malecot then applied this idea to general formulation of the covariances of relatives for inbred populations (see section 2.2.3).

2.2.1.1. The partition of epistatic variation

Cockerham (1954) discussed the case of two alleles per locus with arbitrary gene frequency. The relationship between gene frequencies and genotypic frequencies in a large random mating equlibrium population enabled Cockerham to construct orthogonal scales (contrasts) which give the additive and dominance contributions. He then showed how these scales could be combined to partition the two-factor epistatic variance into additive × additive, additive × dominance, dominance × additive and dominance × dominance components. The analogy with factorial analysis of variance is

quite apparent. With two loci each with two alleles, there are nine possible genotypes and eight single degree of freedom comparisons can be isolated (see Table 2.1). The single epistatic variance term obtained by Fisher (1918) is just the sum of the four epistatic components as indicated above (Cockerham, 1954).

Kempthorne (1954, 1955) extended Cockerham's ideas to the general case of an arbitrary number of loci each with an arbitrary number of alleles. Assuming no linkage, he formulated covariances of relatives in a random mating population (Kempthorne, 1954, 1955) and in a selfing population (Kempthorne, 1956). He also worked out the case for polyploids (Kempthorne, 1957, Ch. 18). On the basis of Malecot's earlier work and using these new results, Cockerham (1954) and Kempthorne (1957) gave a general formula for covariances of relatives from a random mating population in Hardy-Weinberg and linkage equilibria:

$$Cov(X,Y) = 1/2(\Theta+\Theta')\sigma^2_A + \Theta\Theta'\sigma^2_D + 1/4(\Theta+\Theta')^2\sigma^2_{AA} + 1/2(\Theta+\Theta')\Theta\Theta'\sigma^2_{AD} + (\Theta\Theta')^2\sigma^2_{DD} + \cdots , \quad (2.2)$$

where Θ and Θ' are probability measures of identity by descent (see Section 2.2.2). Note that the first two terms are additive and dominance variances and the rest are all possible types of epistatic variance components for the loci under consideration. Cockerham (1956) and Schnell (1963) extended this formula by taking into account the effects of linkage on covariances between the relatives, provided that the initial population is in linkage equilibrium.

Comstock and Robinson (1952) developed the well-known NC experiments I, II and III in order to measure additive genetic variance and

Table 2.1: Factorial analysis of genetic variance for a two-locus model

Source	df	Variance component
Locus A	2	
Additive	1	σ^2_{Aa}
Dominance	1	σ^2_{Da}
Locus B	2	
Additive	1	σ^2_{Ab}
Dominance	1	σ^2_{Db}
A × B	4	
Additive × additive	1	σ^2_{AA}
Additive × dominance	1	σ^2_{AD}
Dominance × additive	1	σ^2_{DA}
Dominance × dominance	1	σ^2_{DD}

estimate the level of dominance. Robinson and Comstock (1955) reported the results of these efforts from their corn experiments.

In connection with these efforts, Horner, Comstock and Robinson (1955) studied non-allelic gene interaction in some symmetric epistatic models and investigated the extent of possible bias in estimates of additive and

dominance variances for the above designs. On the other hand, Gates, Comstock and Robinson (1957) investigated bias due to linkage.

The particular advances made by these two pieces of work are that they developed a general notation and model which can be used to study the covariances of relatives arising from self-fertilization. Let $Cov(k,k-1;n,n')$ be the average covariance between F_k-derived F_n sub-population means and corresponding F_k-derived $F_{n'}$ sub-population means within F_{k-1} sub-populations. Under the assumption of no linkage, Horner *et al.* (1955) found, for the additive model with dominance, that

$$Cov(k,k-1;n,n') = \frac{\sigma_A^2}{2^{k-2}} + \frac{\sigma_D^2}{2^{n+n'-k-2}}, \tag{2.3}$$

where σ_A^2 and σ_D^2 are additive and dominance variances, respectively. The covariance becomes variance when $n = n'$. Thus, $Cov(k,k-1;n,n)$ represents the average variance of F_k-derived F_n sub-populations within the F_{k-1} subpopulation. Horner *et al.* (1955) gave numerical values for $Cov(k,k-1;n,n')$ with different combinations of k, n, n' and genetic models as well as number of loci. Five symmetrical models with or without epistasis were used in their study to assess the possible bias due to epistasis.

Gates *et al.* (1957) generalized the notation such that $Cov(k,k-1;n,n')$ became $Cov(k,k';n,n')$ with the condition $2 \leq k' \leq k \leq n \leq n'$. Since Horner et al.'s (1955) model assumes no linkage, Gates *et al.* (1957) introduced an additional parameter p, the recombination value, into the notation. For an additive-dominance model with linkage, but no epistasis, Gates *et al.* (1957) derived general formulas for $Cov(p;k,1;n,n')$ and $Cov(p;k,k';n,n')$.

In his book entitled "*Biometrical Genetics*", Mather (1949), following the notation and analysis set forth by Fisher *et al.* (1932), presented formulas for some variances and covariances of relatives in early generations of selfing with the assumption of linkage, but no epistasis. Some of these results were generalized by Nelder (1952) for variances in any generation of selfing as well as backcrossing.

Basically, Mather's model uses two parameters, d and h, to describe the three genotypes possible with two alleles per locus (Mather and Jinks, 1971, 1982). Let $2d$ equal the difference between the two homozgyotes, AA and aa, and h equal the deviation of the heterozgyote, Aa, from the homozgyote mean. The effects so defined are then two contrasts among genotypes without reference to any population (Wright and Cockerham, 1986b). This is in contrast to Cockerham's (1954) and Kempthorne's (1957) models in which a random mating equilibrium population provides the reference for defining gene effects. The contributions of these two contrasts to the total genetic variance will be statistically independent so long as the two homozgyotes are equally frequent in the population or families.

Hayman and Mather (1955) extended the Mather model to the case of genes at two loci, each with two alleles A-a and B-b. Nine genotypes are possible in a diploid and eight parameters must be used to give a complete decription of the effects among genotypes. These parameters are referred to as additive effects at both loci, d_a and d_b, dominance effects at both loci, h_a and h_b, and four non-allelic interaction terms, $i_{ab}(d_a \times d_b)$, $j_{ab}(d_a \times h_b)$, $j_{ba}(h_a \times d_b)$ and $l_{ab}(h_a \times h_b)$.

2.2.1.2. Hayman's device

Hayman (1954, 1955) argued that the analogy between a polygenic system and a factorial experiment is only nominal. Such analogy will disappear when consideration shifts from description of the gene action to the problem of measuring it. For example, in a fertilizer trial, the effect of each fertilizer on the yield of the crop is measured (estimated) by taking the differences between the total yields of plots which are subject to different levels of the fertilizer. However, in a large random mating population with genes at many loci, it is not possible to estimate the gene effects at a single locus because the individual genotypes in the population are not usually identifiable. Only estimates of average effects, such as average dominance, over all loci can be made.

Hayman also argued that genetic properties, such as segregation and recombination (linkage), bear no analogy in the factorial experiment. Hayman (1954, 1955) went on to suggest what he called a more natural approach. This consists of scaling the genotype of an individual suitable for algebraic manipulation, expressing the (metrical) phenotype of the individual as a polynomial function of this scalar genotype and then, using Mendel's laws to establish the relation between genotypic representations of parent and progeny, to finally obtain statistics such as means, variances and covariances by the usual algebraic procedures. Kempthorne (1957) called this procedure Hayman's device.

In the Hayman notation, three possible genotypes AA, Aa and aa at locus A-a have phenotypes $c + d_a$, $c + h_a$ and $c - d_a$ respectively, where c is constant, $d_a > 0$ and h_a may take either sign. These genotypes can be

represented by introducing a variable θ (Hayman, 1954) with values 1, 0, and -1 corresponding to genotypes AA, Aa and aa, respectively. Thus, the phenotype is polynomial function of this scalar genotype, i.e.,

$$M(\theta_a) = c_0 + c_1\theta_a + c_2\theta^2{}_a, \tag{2.4}$$

where $M(\theta_a)$ is termed the metric and c_0, c_1 and c_2 can be adjusted to represent measures of gene action. In view of the standard measure of additive and dominance effects, two alternative forms for the metric are

$$M(\theta_a) = m + d_a\theta_a + h_a(\frac{1}{2} - \theta^2{}_a) \text{ and} \tag{2.5}$$

$$M(\theta_a) = m + d_a\theta_a + h_a(1 - \theta^2{}_a). \tag{2.6}$$

The first metric has been called the F_2-metric and the second the F_∞-metric (Van der Veen, 1959). Other notations for gene action, such as Comstock and Robinson's (1952) u and a notation, can be incorporated into Hayman's device. This can be achieved merely by substitution of d and h in Eqn.(2.5-2.6) by u and au, respectively, in the notation of Comstock and Robinson. Van der Veen (1959) showed how the translation of the parameters from one metric into another is made. Mather and Jinks (1971, 1982) have used the F_∞-metric while Kempthorne (1957) used the F_2-metric in discussing the related materials.

The individual with genotype θ will produce gametes containing A and a with respective frequencies, $1/2(1+\theta)$ and $1/2(1-\theta)$. The cross $\theta' \times \theta''$ will produce progenies AA, Aa and aa with frequencies $1/4(1+\theta')(1+\theta'')$, $1/2(1-\theta'\theta'')$ and $1/4(1-\theta')(1-\theta'')$, respectively. This expresses Mendel's law of segregation. The expectations of θ and θ^2 in the progeny are $1/2(\theta'+\theta'')$ and $1/2(1+\theta'\theta'')$, respectively. Thus, for the F_∞-metric, the expected mean phenotype of these progeny is $1/2[d(\theta'+\theta'') + h(1-\theta'\theta'')]$.

In the absence of linkage this device can be extended to the case of genes at many loci with little difficulty (Hayman, 1955). Hayman (1955) gave general formulas for variances and covariances of relatives in a selfing population. In another case, Hayman (1954) showed how the device could be employed to analyze diallel crosses. Jones (1960) extended Hayman's device to include linkage and gave some general results for selfing systems with the presence of linkage and epistasis. However, Jones did not give the recursion relations between generations.

Kempthorne (1957) criticized Hayman's device for the following reason. The utility of the device appears limited to the case of two alleles per locus and the treatment of multiple alleles becomes unnecessarily complicated.

2.2.2. Descent measures and covariances of relatives

The concept of identity by descent was first introduced by Malecot (1948, 1969) to give equivalent, but easier, interpretation of Wright's inbreeding coefficient. Two genes are said to be identical by descent if they are copies of the same ancestral gene. It is this fact that causes the resemblance (or correlation) between relatives and homozygosity in inbred populations (Bulmer, 1980). Genetic relationship of any two related individuals, X and Y, in a population may be characterized by the probability that a gene in X and a gene in Y may both be copies of a gene in one of their common ancestors. Malecot (1948, 1969) called this probability the "*coefficient de parente*". This term has been translated as *coefficient of parentage* by Kempthorne (1957) and *coancestry* or *coefficient of kinship* by Falconer (1981). It is well known (Falconer, 1981) that the

inbreeding coefficient of an individual is equal to the coancestry of its parents.

The concept of identity by descent makes it possible to synthesize Fisher's least-squares procedure of partitioning the genotypic variance and Wright's measure of kinship (Van Aarde, 1975). The general synthesis then provides a framework for obtaining the covariances of relatives under any system of mating.

2.2.2.1. One-locus model

Consider first the case of a single locus where two alleles are carried by individual X and two by its relative Y. The identity relations between genes within two relatives X and Y can be characterized by fifteen mutually exclusive and exhaustive events. The probabilities of these events sum to unity. This has been recognized from somewhat different arguments by Gillois (1964), Harris (1964), Cockerham (1971) and Denniston (1974). Since there is generally no need to distinguish between the paternal and maternal origin of the genes, Jacquard (1974) suggested using a condensed set of nine measures of identity by descent.

For the one-locus model, the covariances of relatives under general inbreeding can be expressed as a linear function of five quadratic components which are defined with reference to a random mating equilibrium population (Gillois, 1964, Harris, 1964, Weir and Cockerham, 1977). The cofficients associated with these population parameters are either sums or differences of the identity by descent (*ibd*) measures and are calculated by algorithms given in several places (e.g., Cockerham, 1971).

In considering the inbreeding system restricted to self-fertilization,

Cockerham (1983) and Cowen (1986) presented a set of *ibd* measures which are required to express the covariances, $C_{tgg'}$, between relatives in two selfed generations, g and g' with the last common ancestor in the ***tth*** generation. With Cockerham's (1983) definitions for the five quadratic components and assumptions of (a) linkage eqilibrium in the initial population, (b) no linkage and (c) no epistasis, both authors found the covariances of relatives from self-fertilization as,

$$C_{tgg'} = (1+F_t)\sigma_A^2 + \frac{(1-F_g)(1-F_{g'})}{(1-F_t)}\sigma_D^2 + (F_g+F_{g'}+2F_t)D1$$
$$+ [F_t + \frac{(F_g-F_t)(F_{g'}-F_t)}{2(1-F_t)}]D2 + [\frac{F_t(1-F_g)(1-F_{g'})}{(1-F_t)}]H^*, \qquad (2.7)$$

where F_x = 1 - $(1/2)^x$ for x = 0, 1, 2, ... It may be noted that both Cockerham and Cowen considered the initial (*zeroth*) generation non-inbred and in linkage equilibrium. In view of standard notation for selfing, consideration takes place only after formation of the F_2 populations (Gates *et al.*, 1957). In this notation (Horner *et al.*, 1955, Gates et al, 1957) the inbreeding coefficient should be F_x = 1 - $(1/2)^{x-2}$ for $x \geq 2$.

In contrast to Cockerham's (1983) model, Wright and Cockerham (1986b) gave an alternative form of the model which extended Mather and Jinks' (1971, 1982) model of two equally frequent alleles at each locus in a selfing population derived from a single F_1 to the case of multiple alleles. They specified a set of quadratic components in terms of the variances of homozygotes instead of the usual additive genetic variances. This results in fewer, but equivalent, quadratic components required to express the

covariances of relatives from selfing (Wright and Cockerham, 1986b). In a recent paper, Cornelius and Van Sanford (1988) compared three equivalent models given by Cornelius and Dudley (1976), by Cockerham (1983) and by Wright and Cockerham (1986b) and showed how one model could be expressed in terms of another model.

Cockerham and Weir (1984) also gave a set of *ibd* measures for the case of mixed self and random mating, and provided a formula for the covariances of relatives derived from this case. Since several economically important crop species have such mating systems, it was natural for Wright and Cockerham (1985, 1986a) to study selection responses by using covariance techniques.

2.2.2.2. Two-locus model

In a single locus case, any individual in a population has only a pair of genes. The only measure of inbreeding in the population is Wright's one locus inbreeding coefficient, F, the probability that the two genes at the locus in the individual are identical by descent (Malecot, 1948). Shifting to the two-locus case where the individual has four genes, A_1, B_1, A_2 and B_2, Cockerham and Weir (1973) and Weir and Cockerham (1973) found that there are six pairs of genes, and 15 possible arrangements in gametes of the initial individuals from which A_1, B_1, A_2 and B_2 are descended. This led them to define a set of *descent measures* for two linked loci. They expanded Malecot's (1948) concept of identity by descent by introducing *equivalence by descent*, a probability measure that a pair of genes are both descended from genes in one initial gamete. Thus, allelic genes which are equivalent by descent are also identical by descent (Weir and Cockerham, 1977). With the

assumption that each locus is equally inbred, Cockerham and Weir (1973) recognized that the 15 possible arrangements reduce to nine. Since the probabilities, or equivalence by descent measures, of the nine arrangements sum to one, only eight descent measures are necessary. General procedures for evaluating the descent measures in any situation have been developed by Cockerham and Weir (1973) and can be applied to a specific mating system, for example, selfing (Weir and Cockerham, 1973).

The eight descent measures and their probability definitions for an inbred individual with genes A_1B_1/A_2B_2 are given in Table 2.2. The first measure, F_1, is Wright's (1921) one-locus inbreeding coefficient. F_{11} is the two-locus inbreeding coefficient which was first introduced by Haldane (1949) and has been investigated extensively by several authors (e.g., Schnell, 1961; Narain, 1965; Shikata, 1965; Weir and Cockerham, 1969). The difference between the two-locus inbreeding coefficient and the squared one-locus inbreeding coefficient is defined as identity disequilibrium (Weir and Cockerham, 1969) and becomes zero in the absence of linkage. The remaining six measures are the new features of Cockerham and Weir's two-locus descent measure theory. For a review on this theory, see Cockerham and Weir (1977). These four-gene descent measures have been used to provide means and variances for a two-locus quantitative model of gene effects in a full generality (Weir and Cockerham, 1977).

To derive the covariances between two relatives, X and Y, with respect to the two-locus model, a set of eight-gene descent measures is required. Such measures would give the probabilities of all the ways in which eight genes shared by two individuals could be located on one to eight initial

Table 2.2: A complete set of *ibd* measures of an inbred individual for a two-locus model (from Weir and Cockerham, 1977)

Measure	Definition†
F_1	$\frac{1}{2}[\,P(A_1 \equiv A_2) + P(B_1 \equiv B_2)\,]$
F^1	$\frac{1}{2}[\,P(A_1 \equiv B_1) + P(A_2 \equiv B_2)\,]$
${}_1F$	$\frac{1}{2}[\,P(A_1 \equiv B_2) + P(A_2 \equiv B_1)\,]$
${}_1F_1^1$	$\frac{1}{4}[\,P(A_1 \equiv B_1 \equiv B_2) + P(A_2 \equiv B_1 \equiv B_2) + P(A_1 \equiv A_2 \equiv B_1) + P(A_1 \equiv A_2 \equiv B_2)\,]$
F_{11}	$P(\,A_1 \equiv A_2 \; and \; B_1 \equiv B_2\,)$
F^{11}	$P(\,A_1 \equiv B_1 \; and \; A_2 \equiv B_2\,)$
${}_{11}F$	$P(\,A_1 \equiv B_2 \; and \; A_2 \equiv B_1\,)$
F_{11}^{11}	$P(\,A_1 \equiv B_1 \equiv A_2 \equiv B_2\,)$

† $\equiv$ denotes identity or equivalence by descent.

gametes. Gallais (1970, 1974) identified components of expected covariances between the two relatives for a two-locus model, but has given no general indication of how to evaluate the coefficients associated with the components.

Because of the unwieldy expression for the variances and the even more complicated situation expected for the covariances under a completely general model, Weir and Cockerham (1977) suggested the use of approximation approaches. This has been done by several authors. By assuming that the initial population is non-inbred and in linkage equilibrium, Schnell (1963)

used his own concept of inbreeding function (Schnell, 1961) to investigate effects of linkage on the partition of epistatic variation in the covariances of relatives. This inbreeding function generalizes Wright's inbreeding coefficient to the multi-locus case. Schnell's work, therefore, serves as an extension of Cockerham-Kempthorne's general formula without linkage [see Eqn.(2.2)].

Instead of working with general inbreeding where the explicit expressions for the covariances become unwieldy under complex pedigrees, Cockerham (1984b) and Cockerham and Matzinger (1985) concentrated on the covariances of relatives from selfing and obtained two-locus descent measures for additive $\times$ additive epistatic variance in the covariances. In addition, Cockerham (1984a) and Cockerham and Tachida (1988) took both linkage and genetic drift into account in expressing the covariances.

Wright (1987, 1988) assumed linkage eqilibrium in the initial population and no linkage, and gave the two-locus covariances under general inbreeding as well as selfing for both the conventional model (Weir and Cockerham, 1977) and the homozygote based model (Wright and Cockerham, 1986). For this latter model, only 12 quadratic components are necessary in selfing (Wright, 1988).

With assumptions of linkage equilibrium and no inbreeding, Denniston (1975) and Van Aarde (1975) evaluated the various descent measures which are required for expressing the covariances between non-inbred relatives from a linkage equilibrium population. The two-locus variances and covariances in such cases involve only the five variance components, σ^2_A, σ^2_D, σ^2_{AA}, σ^2_{AD} and σ^2_{DD}.

2.2.3. Estimation of genetic variances and covariances

To estimate genetic variances and covariances of relatives, some system of mating, the mating design, is used to generate sets of relatives raised in one or more environments, the environment design (Cockerham, 1963).

In an extensive investigation on corn quantitative genetics, Comstock and Robinson (1952) described three experiments for estimating genetic variances and covariances. These experiments have become known as North Carolina Experiments I, II and III.

The mating designs involved in these experiments have a statistical analogy. For example, Experiment I is often called a hierarchical (nested) design, while Experiment II is a factorial design. Cockerham (1963) denoted I as A/B and II as AB, where A and B represent the two sets of parents (males and females). In a diallel mating design, only *one* set of individuals is used for both male and female parents. Cockerham denoted diallel designs as AA.

The diallel mating design produces a series of full-sibs and half-sibs. Estimation of the covariances among these relatives has been discussed by Cockerham (1963) and Hallauer and Miranda (1981). The analysis and interpretation have been given by Griffing (1956) and Kempthorne (1957) for the case where parents are selected at random from a random mating population, and by Hayman (1954) and Jinks (1954), as well as Mather and Jinks (1971, 1982), for the case where parents are inbreds. A critical review by Baker (1978) centers around some assumptions required for valid statistical analysis and genetic interpretation of diallel mating designs.

Experiment III of Comstock and Robinson is the most efficient of the

three experiments (Bulmer, 1980). In this experiment, two homozygous parents are crossed and the resulting F_2 plants are backcrossed to each of the parents. Kearsey and Jinks (1968) expanded this concept to develop a new mating design, called a *triple test cross*. In a triple test cross, a random sample of the F_2 plants is backcrossed to each of the parents as well as to the F_1. Kearsey and Jinks, based on the triple test cross, constructed an unambiguous test for epistasis.

Although many mating designs have become available for cross-fertilized crops there is only one type, nested mating design, for the segregating generations of self-fertilized crops (Wricke and Weber, 1986). The nested mating design is a consequence of using the natural reproduction mechanism of self-fertilization (Cockerham, 1963). The description and analysis of this design has been the subject of papers by Horner *et al.* (1955), Horner and Weber (1956) and Gates *et al.* (1960). Among other authors are Cockerham (1963) and Weber (1982) who used different notations.

Estimation of genetic variances and covariances of relatives developed from the experiments described above can proceed in different ways. The traditional least-squares method of Fisher (1918) involves calculating the mean squares for each effect and equating each mean square to its expected linear function of variance components. The phenotypic covariances of relatives, derived from the design variances and covariances, can be interpreted in terms of various genetic variance components, depending on the models, plus environmental variance. Thus, the genetic variance components can be estimated by the least-squares procedure.

In many cases, this procedure gives simple and efficient estimates of

genetic variance components, as described in several places (e.g., Kempthorne, 1957; Cockerham, 1963; and Mather and Jinks 1971, 1982). However, it sometimes yields negative estimates of variance components. Moreover, this method gives equal weights to mean squares which have different precision and which may be correlated.

Alternative methods are necessary to overcome these deficiencies. Hayman (1960) suggested use of maximum likelihood (ML). The ML method takes into account both unequal variances and correlation between the estimates and always gives positive estimates. In the past, the ML method has not been widely used because of computational and theoretical difficulties. Harville (1977) discussed some recent theoretical developments on the ML method and has suggested several iterative algorithms which can be used for computing the ML estimates of variance components.

The precision of any estimate of variance or covariance is usually indicated by its standard error. A larger standard error suggests a less precise estimate. In planning an experiment to estimate genetic variances and covariances of relatives, one likes to know how many replications, families, sub-families, etc., are needed to achieve a given degree of precision; and to achieve the greatest possible precision, within the limitations imposed by the available facilities, one needs to know the best design for the experiment. These problems were discussed in some detail by Cockerham (1963).

Robertson (1959) discussed the optimum experimental design for estimation of genetic variances and covariances in terms of intraclass correlation and its standard error. He concluded that family sizes in the

order of 2 to 3 are extremely inefficient, and within a single grouping and the intraclass correlation t, the optimum family size is approximately 1/t. In Robertson's opinion, when one has little *a priori* knowledge of the magnitude of t, he should be concerned with a range of t values from 0.01 to 0.10.

In estimating genetic variance components in the covariances, Sentz (1971), Cockerham (1983) and Cornelius and Van Sanford (1988) recognized that the estimate of dominance variance has a larger standard error than that of additive variance because the coefficients of different components are highly correlated for many sets of relatives (collinearity problem). Bridges and Knapp (1987) argued that the probability of negative estimates is generally lower for additive variance than for dominance variance. In terms of experimental design, they found that such probability is lower for the factorial mating design than for the nested mating design.

2.3. Methods for detecting linkage

Methods for detecting the presence of linkage have been devised with reference to particular populations. If a large random mating population is assumed in linkage equilibrium, then linkage can increase the covariances of sibs (full and half sibs) through an increase in the components due to non-allelic interaction (epistasis); the closer the linkage the greater the increase. With no epistasis, there is no effect of linkage on the covariances of relatives (Cockerham, 1956) in a large random mating population. The test for linkage stemming from this argument will only be applicable to cross-fertilized species.

On the other hand, linkage disequilibrium exists in an F_2 population derived from a cross of two inbred lines. Mather (1949) and Gates *et al.* (1957, 1960) recognized that, in the absence of epistasis, the mean is not affected by linkage. It was natural, then, for them to suggest use of the second order statistics, variances and covariances, to detect linkage.

2.3.1. Statistics of rank for detecting linkage

2.3.1.1. Theory

Mather (1949) presented the first test for linkage in studying the inheritance of quantitative traits in inbred populations. Consider the simplest case where there are two loci, A-a and B-b, with recombination value p between the two loci and no epistasis between the two loci. Let values of three genotypes at each locus be d, h and $-d$, then the genetic variance in the F_2 is

$$V_{1F2} = \frac{D}{2} + \frac{H}{4},$$

with

$$D = [d_a^2 + d_b^2 \pm 2(1-2p)d_a d_b],$$
$$H = [h_a^2 + h_b^2 + 2(1-2p)^2 2h_a h_b],$$

where $\pm$ should be read as coupling (+) or repulsion (-) linkage in the F_1. Mather also gave the expected values of three other statistics, variance among the F_3 family means, V_{1F3}, covariance between the F_2 and the F_3, W_{1F23} and the mean variance within the F_3 families, V_{2F3}. He found from these expectations that the expressions for D and H given by V_{1F2}, W_{1F23} and V_{1F3} are the same, but are different from those given by V_{2F3}. More

specifically, $(1-2p)$ and $(1-2p)^2$ of D and H in V_{1F2}, W_{1F23} and V_{1F3} are replaced by $(1-2p)^2$ and $(1-2p)^2(1-2p+p^2)$ of D and H in V_{2F3}.

This difference provides the basis on which a test for linkage is established. Mather (1949) argued that variation measured by V_{1F2}, W_{1F23} and V_{1F3} reflects the consequence of one round of recombination at gametogenesis in F_1 and, thus, they are all first rank statistics. V_{2F3} reflects the effect of recombination at gametogenesis in F_2 as well as that in F_1 and, thus, it is a second rank statistic. It is apparent that the definitions in D and H have changed with the rank of the statistics. The test of linkage is then a test of heterogeneity of D and H over the statistics of different rank. In the absence of epistasis, D and H are homogenous over V_{1F2}, W_{1F23} and V_{1F3}, but will change in V_{2F3} where linkage is acting (Hayman and Mather, 1955).

2.3.1.2. Results

Mather (1949) described an experiment on the inheritance of ear conformation in a barley cross, Spratt × Goldthorpe. The experiment was set out in five randomized blocks with 117 plots per block. Of these, 100 were assigned to 100 different F_2-derived F_3 families selected from 170 F2 plants, 10 to the F_2, 3 to F_1, and 2 to each parental variety. V_{1F2} and V_{2F3} were estimated from the usual variances among and within the F_2-derived F_3 families. V_{1F3} and W_{1F23} were calculated using the deviations from block means for each family. After estimating additive and dominance variances, 2 df are left; 1 df is assigned to test for linkage over ranks and 1 to the residual interaction. Two environmental variances, error variances among, Eb, and within, Ew, the families were also calculated directly from each block of

the experiemnt. Some 37% of total variation after fitting additive and dominance variances was explained by the effects of linkage, but only 0.1% by the residual interaction (Mather, 1949). Mather indicated that the value of V_{2F3} was less than one half of V_{1F2} and concluded that coupling linkage was present.

Linkage studies on number of sternopleural chaetae in *Drosophila melanogaster* were conducted by Cooke *et al.* (1962) and Cooke and Mather (1962). The four statistics, V_{1F2}, W_{1S23}, V_{1S3}, and V_{2S3}, were calculated from this experiment, where S in the three statistics stands for sib-mating. In analyzing these variances and covariances, Cooke *et al.* (1962) showed a significant (P = 0.05) effect due to linkage. In addition, Cooke *et al.* (1962) treated both the sexes and the descendants of the two reciprocal crosses separately so as to give a means of detecting heterogeneity of linkage effects with 3 df. The result of this test was also significant at P = 0.05.

Jinks and Pooni (1982) and Pooni and Jinks (1986) raised a question of how to estimate the true additive genetic variance, ${}_{D}V_{F\infty}$, in the presence of linkage. In their definition,

$$ {}_{D}V_{F\infty} = \sum_{r=1}^{\infty} (\frac{1}{2})^{r} D_{r}, \qquad (2.8) $$

where, for two loci, the additive variance of rank r, $D_r = d_a^2 + d_b^2 \pm 2(1-2p)^r d_a d_b$. Because the coefficient of D_r follows the geometric series $(1/2)^r$, successive terms make a rapidly decreasing contribution to total variance. It is of little error to make the following equality: $D_2 = D_3 = \cdots = D_\infty$. Therefore, $1/2D_1 + 1/2D_2$, or more precisely, $1/2D_1 + 1/4D_2 + 1/4D_3$ may be acceptable approximations in practice to ${}_{D}V_{F\infty}$.

The additive genetic variances of different ranks, D_1, D_2, D_3 . . . , can be estimated from a pedigree inbreeding design of self-fertilization or from sib-mating. For example, F_3 families from F_2 plants, or S3 families by sib-mating F_2 pairs at random, will give estimates of D_1 and D_2 from their expectations:

Among families	$V_{1F3} = 1/2D_1 + 1/16H_1$	$V_{1S3} = 1/4D_1 + 1/16H_1$
Within families	$V_{2F3} = 1/4D_2 + 1/8H_2 + Ew$	$V_{2S3} = 1/4D_2 + 3/16H_2 + Ew$

Thus, $D_1=4(V_{1F3}-V_{1S3})$, $D_2=12V_{2F3}-8V_{2S3}-4Ew$ and $1/2D_1+1/2D_2=2V_{1F3}+6V_{2F3}-2V_{1S3}-4V_{2S3}-2Ew$, where estimates of Ew can be obtained by including the parental P1 and P2 as well as their F_1 families in the experiment.

Jinks and Pooni (1982) estimated D_1, D_2 and D3 from the F_3 and F_4 of a pedigree selfing series from the cross between varieties 1 and 5 of *Nicotiana rustica*. A previous study showed that flowering time and plant height both had a significant repulsion linkage. For both traits, the estimated D_1 was significantly smaller than D_2 and D_3, but D_2 and D_3 did not differ significantly. The predicted value, ${}_DV_{F\infty}$, for flowering time, using $1/2D_1+1/2D_2$, was 44.5, while the observed value ${}_DV_{F\infty}$, approximated by the estimates of pedigree inbreeding in F_6 and by highly inbred families of single seed descent in F_{15}, were equal to 39.4 and 38.4, respectively.

2.3.2. Gates' test for linkage

2.3.2.1. Theory

As an alternative to the rank statistics method, Gates *et al.* (1960) proposed a test for linkage using general forms of $Cov(k,k';n,n')$. The assumptions made in Gates' test are:

(1) no epistasis;

(2) two equally frequent alleles per locus;

(3) no mutation;

(4) no selection;

(5) no cytoplasmic or maternal effects.

In Gates' test, approximate expectations of the covariances of relatives were developed [The exact expectations were given in an earlier paper by Gates *et al.*(1957)]. The six coefficients derived from the approximate expectations serve as independent variables in a multiple regression of a series of covariances. The relative importance of linkage can be determined by analysis of variance in the multiple regression. The subsequent regression coefficients provide estimates of additive and dominance variances, as well as four quadratic functions associated with linkage.

Let $Cov(k,1;n,n')$ be the genotypic covariance between F_k-derived F_n progeny means and F_k-derived $F_{n'}$ progeny means within F_1. Then,

$$Cov(k,1;n,n') \simeq \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6. \qquad (2.9)$$

Values of X_1 to X_6 can be determined as simple functions of k, n and n'. They were given in Gates *et al.* (1960). The general form $Cov(k,k';n,n')$ can be found by the difference, $Cov(k,1;n,n')-Cov(k'1;n,n')$.

β_1 and β_2 are the usual additive and dominance variance. The four additional terms, β_3 to β_6, arise because of linkage and they should vanish when there is no linkage. The expressions for the odd number terms, β_3 and β_5, are related in form to the expression for the additive variance, β_1, while those for β_4 and β_6 are related to that for dominance variance, β_2. For a pair of linked loci, *A-a* and *B-b*, with recombination value p, the explicit expressions of β_1 to β_6 in Gates' test are:

$$\begin{aligned}
\beta_1 &= [d_a^2 + d_b^2 \pm 2(1-2p)d_a d_b]/2 \\
\beta_2 &= [h_a^2 + h_b^2 + 2(1-2p)^2 h_a h_b]/4 \\
\beta_3 &= p[\pm(1-2p)d_a d_b] \\
\beta_4 &= p(1-p)[(1-2p)^2 h_a h_b] \\
\beta_5 &= p^2[\pm(1-2p)d_a d_b] \\
\beta_6 &= 2p^2(1-p)^2[(1-2p)^2 h_a h_b],
\end{aligned} \tag{2.10}$$

where d and h are the measures of additive and dominance effects and $\pm$ in β_1, β_3 and β_5 should be read as coupling (+) and repulsion (−) linkages respectively. Thus, a positive sign in estimates of β_3 and β_5 indicates a preponderance of coupling linkages of genes conditioning a quantitative trait, while a negative sign in the same estimates indicates a preponderance of repulsion linkages. The sign of β_4 and β_6 is independent of linkage phases, but not of dominance effects, h_a and h_b. If the majority of the more favourable alleles at the linked loci are dominant to their less favourable alleles, or vice versa (i.e., h_a and h_b have the same sign) then estimates of β_4 and β_6 will be positive. With a mixed sign among the alleles at the linked loci (i.e., h_a and h_b have different sign), negative estimates of β_4 and β_6 may be achieved.

2.3.2.2. Results

Gates *et al.* (1960) used Gates' test for linkage to study eight quantitative traits in a soybean [*Glycine max* (L.) Merrill] cross, Adams × Hawkeye. In order to estimate appropriate variances and covariances of relatives, they utilized the following structure of data (Fig. 2.1). Ninety-four F_2 plants were randomly selected from about 300 F_2 individuals from the cross. This gives 94 F_3 families. In each of these F_3 families, two random plants were chosen with sufficient seed for the next generation. These produced 94 pairs of F_4 families. Two random plants were chosen from a random member of each pair of F_4 families. These produced 94 pairs of F_5 families. This pattern was then repeated up to F_7. These different generations were grown in different years.

According to Horner and Weber (1956) and Gates *et al.* (1960), analysis of this data structure resulted in nine sample variances (through the usual ANOVA procedures) and 21 sample covariances. Gates *et al.* (1960) calculated these 30 variances and covariances for eight traits, flowering time, period from flowering to maturity, maturity date, height, lodging, seed yield, seed weight and percent oil. Multiple regression of these 30 covariances on the 30 sets of X_1 to X_6 was performed. The analysis provided a test of significance for effects due to linkage as well as additive and dominance effects. The relative importance of linkage was indicated by the test of significance for reduction in sums of squares due to regression on X_3 to X_6 excluding X_1 and X_2, X_4 to X_6 excluding X_1 to X_3, X_5 and X_6 excluding X_1 to X_4 and X_6 excluding X_1 to X_5.

Results from this analysis (Gates *et al.*, 1960) indicated that linkage

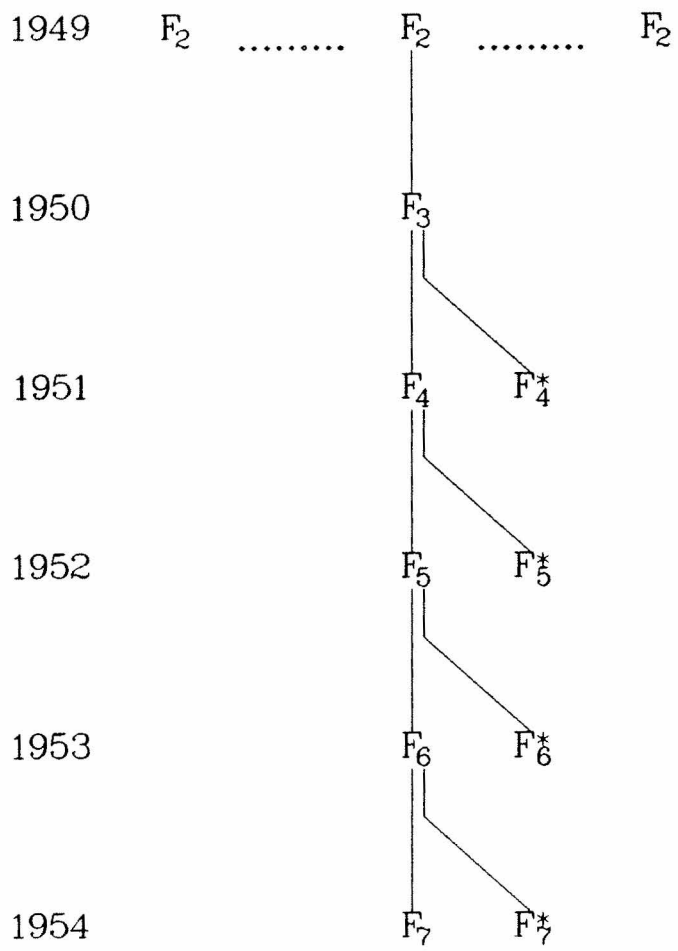

Figure 2.1: The hierarchical structure used by Gates *et al.* (1960).

was important only for flowering time, height and yield. Furthermore, estimates of β_3 for height and yield were negative, suggesting the presence of repulsion linkages, while the estimates of β_3 for flowering time was positive, suggesting the presence of coupling linkage.

Croissant and Torrie (1971) studied two soybean crosses, Norchief × Clark (N × C) and Norchief × Harosoy (N × H), for seven quantitative traits (the same traits as those of Gates *et al.* except for percent oil). Gates' test was utilized to detect nonadditive effects and linkage.

For each cross a random sample of 49 F_2 plants was grown to establish F_3 families. The procedure of Gates *et al.* (1960) was used to create 49 pairs of families in F_4 to F_7, but the four generations (F_4 to F_7) were planted in the same year for each of two years. In each year two replicates of a split-plot design for both crosses were grown; the 49 F_2 families formed whole plots treatments, and the eight progenies, subplots treatments. This experimental design should allow for unbiased estimation of genetic parameters which are otherwise affected by the genotype-environmental interactions (Croissant and Torrie, 1971).

The analysis of all the four generations provided a set of 20 sample variances and covariances. The expectations of these statistics were calculated according to Horner and Weber (1956). The relative importance of additive, dominance, and four linkage components of genetic variance was determined by the analysis of multiple regression of the 20 covariances on the the 20 sets of X_1 to X_6, as described by Gates *et al.* (1960). From these analyses, Croissant and Torrie (1971) found that additive variance was the principal source of genetic variance and covariance for all seven traits in both crosses.

Significant dominance effects were found for maturity in N × C, flowering time and lodging in N × H, and for height and seed weight in both crosses. Linkage was of importance only for flowering time, height, seed weight and lodging in N × H, and for height and seed weight in N × C. Except for seed weight in N × H, coupling linkages were suggested for all the traits where linkage was detected.

VandeLogt (1978) and VandeLogt *et al.* (1984), using Gates' (1960) test for linkage, evaluated components of genetic variance and the effects of linkage for five quantitative traits, heading date, plant height, kernel brightness, test weight and grain yield, in two barley (*Hordeum vulgare* L.) crosses, Bonanza × H279-70-1-4 and Larker × H279-70-1-4. Their procedures for developing the data structure and analyzing data were exactly the same as those of Croissant and Torrie (1971). Results from their studies indicated that additive genetic variance was an important component of the observed variances and covariances for all traits in both crosses. Dominance variance was important for heading date, kernel brightness, test weight and grain yield, but may have been inflated by linkage effects, as indicated by the authors. Their results also showed a coupling linkage for grain yield in both crosses and for heading date in Larker × H279-70-1-4. Since the three parents were elite barley genotypes, the significant coupling linkages were expected (VandeLogt *et al.*, 1984).

2.3.3. Cockerham's method

In discussing the important question of whether a serious error occurs from ignoring linkage encountered in the covariances of relatives, Cockerham (1984b) gave a general formulation of two-locus coancestries using two-locus identity by descent measures (Cockerham and Weir, 1973). Under assumptions of linkage equilibrium in the ancestral generation, these measures provide coefficients for the additive by additive component in the variances and covariances of relatives for a quantitative trait. He, then, calculated the relative effect of linkage from a ratio of two-locus coancestry, $\bar{\theta}$, to the product of one-locus coancestries, θ^2, $\bar{\theta}/\theta^2$. For example, the relative effect for relatives following self-fertilization is

$$\frac{\eta_{tgg'}}{\theta^2_{tgg'}} \tag{2.11}$$

where η is identity disequilibrium, usually expressed as the difference, $\bar{\theta}$ - θ^2, and should be greater than zero when linkage is present (Weir and Cockerham, 1969). Cockerham's calculation vividly demonstrates the effect of increased inbreeding of the ancestor on reducing the relative effect of linkage in the covariances of descendants. Cockerham concluded that only in the first two generations and with very tight linkage is the relative effect appreciable. In considering the estimation of genetic components of variances and covariance with linkage, he suggested use of relatives with as little inbreeding as possible.

2.4. Computer simulation in genetic research

Scheinberg (1968) outlined four aspects where computer technology can be used in the field of genetics:

1. To design efficient field and laboratory experiments;
2. To analyze the data from the experiment by the computer coded statistical procedures;
3. To solve problems involving complex numerical computation of formulas and equations which cannot be generalized; and
4. To simulate the real and/or model biological system according to the available genetic theories.

Computer simulation (the fourth aspect) has been used in quantitative genetic research to narrow the apparent gap between theoreticians and experimentalists (Fraser and Burnell, 1970; Mulitze, 1983). In a review paper, Geldermann and Gundel (1979) discussed some advantages of computer simulation in relation to real experiments. These advantages include less expense in material and time, better control of more factors and/or exclusion of undesirable influences, total exclusion of systematic errors, repeatable under equal conditions and easier interpretation of results. In their opinion, simulation studies can be used to:

(a) perfect the experiment as a way of instruction;

(b) build and test models;

(c) examine the robustness of a model or a statistical method;

(d) gain theoretical statements from existing models;

(e) help plan an experiment; and

(f) stimulate new experimental or theoretical investigations.

There are generally two types of the computer simulation, deterministic and stochastic (probabilistic), depending on whether or not they are directly concerned with the behaviour and outcome of a random process. According to Mulitze (1983), a deterministic simulation involves specifying probability distributions (e.g., mating probabilities) and computing the results after many generations, while a stochastic simulation in this case involves the repetition of sampling gametes through generations of gametogenesis. Another use of stochastic simulation in quantitative genetic research is to add the random environmental variation to genotypic value so that experimental designs used for estimation of genetic parameters can be evaluated by means of computer simulation (Mulitze, 1983).

The question of whether a deterministic or stochastic simulation should be used in a particular study has been discussed by Crosby (1973). The deterministic simulation is designed to give a complete description of probability distributions such as the mating probability distribution. Any expected values for describing a 'population' can be derived according to the distribution. However, the number and types of matings increase geometrically and the genetic complexity quickly rules out the possibility of a complete description of the probability distributions (Mulitze, 1983). The alternative approach is to get a 'random sample' generated by the use of random numbers. Thus, the expected probability distributions can be estimated from this sample. In addition, simulation of the real biological system requires stochastic models (Crosby, 1973) because both fixable (genetic) and unfixable (environmental) components exist in the system.

Technology on computer simulation, particularly stochastic simulation,

has been a subject of many books and papers, including those by Naylor *et al.* (1968) and Kennedy and Gentle (1980) on computer simulation techniques, and by Gill (1965), Fraser and Burnell (1970) and Crosby (1973) on genetic simulation. A comprehensive account on this was given by Mulitze (1983) who also gave some examples of simulation of genetic systems. A few papers primarily concerned with linkage studies of quantitative traits will be briefly reviewed here.

Bliss and Gates (1968), using a completely additive model with 40 loci, investigated the effect of linked loci on additive genetic variance. They first calculated the expected additive variance for an F_2 population. Then estimates of additive variance were obtained from simulation via the random walk method with 10 replicate computer runs and a 95% confidence interval was constructed from each estimate using the usual χ^2 criterion. They found that a large number of confidence intervals failed to include the expected variance, with poorer fits associated with tighter linkages. They also observed that linkage biased the estimates of additive variance negatively more frequently than positively, although the magnitude of positive bias was greater.

An extensive computer simulation was conducted by Robertson (1970, 1977) to study the effects of linkage on artificial selection. The limits of selection were evaluated by the ratio of the expected value with free recombination (L_f) to that with no recombination (L_0), L_f/L_0. He found that, for a given initial amount of genetic variance controlling a quantitative trait, as the number of loci affecting the trait increased the selection process quickly became independent of the number of loci on a chromosome of a

given length and of their initial gene frequency. He concluded that the effect of recombination on selection response does not appear in the early generations. However, his conclusion was based on the assumption of linkage equilibrium in the initial population.

A paper by Bailey and Comstock (1976) was extensively reviewed by Mulitze (1983). This paper and an earlier paper by Qureshi and Kempthorne (1968) both investigated the impact of linkage on gene fixation probabilities. The latter paper was also concerned with how linkage affects genetic variance during selection. Thompson and Kaiser (1979) used computer simulation to test the power and limitations of the polygene location method proposed by Thoday (1961).

3. Evaluation of Gates' Test for Linkage

3.1. Theoretical evaluation

3.1.1. Assumptions and notation

Gates' test for linkage requires the following assumptions (cf. Section 2.3): (1) normal diploid behaviour at meiosis; (2) no maternal or cytoplasmic effects; (3) no position effects; (4) no multiple alleles; (5) no mutation; (6) no selection; (7) no non-allelic interaction (epistasis); (8) two equally frequent alleles (gene frequency = 0.5). In order to facilitate evaluation of linkage with the presence of epistasis, the assumption of no epistasis in Gates' test will be removed. A general expression for two-locus variance and covariance of relatives from selfing with the presence of both linkage and epistasis will be derived under assumptions set forth above (except for assumption (7)).

Much of the notation employed throughout this study is consistent with that of Horner *et al.* (1955) and Gates *et al.* (1957). However, with regard to genic action, the notation set forth by Mather and Jinks (1971, 1982) will be used to derive the expected variances and covariances. For a one-locus model, Horner *et al.* (1955) and Gates *et al.* (1957) used symbols $2u$, $u + au$ and 0[1] to represent coded genotypic values for genotypes AA, Aa and aa,

[1]This notation was set forth by Comstock and Robinson (1952). The magnitude of a measures the degree of dominance in the action of any pair of genes at a locus. See Horner *et al.* (1955) for full description of qualitative classification of dominance.

respectively. The symbols u and a are equivalent to d and h/d, respectively, in the notation of Mather and Jinks.

Consider the situation where reproduction in all generations after a cross of two homozygous lines is exclusively by self-fertilization. Conceptually, plants in any generation of selfing can be subdivided into one or more groups or subpopulations such that each subpopulation traces to a different individual in any preceding (ancestoral) generation. For example, plants in the F_4 can be subdivided into one subpopulation tracing to the F_1, as many subpopulations tracing to individuals in the F_2 as there are F_2 plants, as many subpopulations tracing to the F_3 as there are F_3 plants, and each F_4 plant represents a subpopulation in the F_4.

With hierarchical mating systems, which are a natural consequence of self-fertilization, the total genetic variance in the F_4 can be subdivided into meaningful components representing (1) the average variance among F_4 plants within F_3 subpopulations, (2) the average variance among F_3 subpopulation means within F_2 subpopulations, and (3) the variance among F_2 subpopulation means. In terms of Gates' notation, this statement becomes,

$$Cov(4,1;4,4) = Cov(4,3;4,4) + Cov(3,2;4,4) + Cov(2,1;4,4).$$

Similarly, the genetic variance of means of F_3-derived F_4 subpopulations is

$$Cov(3,1;4,4) = Cov(3,2;4,4) + Cov(2,1;4,4).$$

Thus,

$$Cov(4,1;4,4) = Cov(4,3;4,4) + Cov(3,1;4,4)\,. \tag{3.1}$$

The hierarchical subdivision of genetic variance can be extended to any

generation by reference to the descendant itself or to a common ancestor in the preceding generation. Such extension is also applicable to the general case of covariance between relatives of two descendant generations tracing to a common ancestor in a preceding generation. In symbols,

$$Cov(k,1;n,n') = Cov(k,k-1;n,n') + Cov(k-1,k-2;n,n') + \cdots + Cov(2,1;n,n')$$

$$= \sum_{t=2}^{k} (t,t-1;n,n') . \tag{3.2}$$

Analogous to Eqn.(3.1),

$$Cov(k,1;n,n') = Cov(k,k-1;n,n') + Cov(k-1,1;n,n') . \tag{3.3}$$

Transposing terms,

$$Cov(k,k-1;n,n') = Cov(k,1;n,n') - Cov(k-1,1;n,n') , \tag{3.4}$$

where $2 \leq k \leq n \leq n'$ which must hold for Eqn.(3.2-3.4) since there have only been k - 2 generations of selfing (i.e., after formation of the F_2). When $n = n'$, $Cov(k,1;n,n')$ represents the variance among means of descendants in the F_n generation tracing to an ancestor in the F_k. When $k = n = n'$, $Cov(k,1;k,k)$ represents the total variance in the F_k generation.

It may be noted that equivalence to the notation of Mather and Jinks (1971, 1982) is:

$$\begin{aligned} Cov(k,k-1;n,n') &= W_{(k-1)Fnn'} ; \\ Cov(k,k-1;n,n) &= V_{(k-1)Fn} , \end{aligned} \tag{3.5}$$

where k - 1 is known as the rank of statistics, in the notation of Mather and Jinks. Rank is, as seen in Section 2.3, of particular importance in relation to linkage.

The more general expression of $Cov(k,k-1;n,n')$ is $Cov(k,k';n,n')$

merely by substitution of k - 1 by k'. This implies that there is no hierarchical classification (or grouping) made from the $F_{k'}$ to the F_k. Unless $k' = k - 1$, k' no longer represents Mather and Jinks' (1971, 1982) rank, but is a sum of ranks from k' to $k-1$. Mather and Jinks' rank may be called *simple* rank, while the sum of ranks is a *compound* rank. $Cov(k,1;n,n')$ was used for deriving Gates' test for linkage (Gates *et al.*, 1960). An analogous expression to Eqn.(3.3) is as follows,

$$Cov(k,k';n,n') = Cov(k,1;n,n') - Cov(k',1;n,n'), \tag{3.6}$$

where $1 \leq k' < k \leq n \leq n'$ for any values of k and k'. This demonstrates a slight deviation from the condition given by Gates *et al.* (1957) where $2 \leq k' \leq k \leq n \leq n'$.

With regard to linkage parameters, the symbol p will be referred to as the recombination value. The following are some simple functions of p:

$$\begin{aligned} q &= 1 - p; \\ \lambda &= 1 - 2p; \\ \xi &= p^2 + q^2 = 1 - 2pq = 1 - 2p + 2p^2. \end{aligned} \tag{3.7}$$

where λ is called as linkage value (Schnell, 1961) and equal to zero with no linkage($p = 0.5$) and to unity with complete linkage ($p = 0$), and $\xi/2$ is equal to frequency of double heterozygotes in F_2 (Bennett, 1954). Another relation is apparent: 2ξ - 1 = λ^2.

3.1.2. Genotypic model

3.1.2.1. Extension of Hayman's device

Consider two alleles A and a at locus A and the two alleles B and b at locus B, and suppose that p is the recombination value between these two loci. In the absence of linkage, nine possible genotypes can be represented by combinations of two variables, θ_a and θ_b, each taking the values 1, 0 and -1. In the presence of linkage, double heterozygotes, $AaBb$, have to be separated into coupling (AB/ab) or repulsion (Ab/aB) phases. For two linked loci, linkage only influences the frequencies of types of gametes produced by the double heterozygotes. It is necessary to distinguish the double heterozygotes from the remaining eight genotypes. Let $\delta_{\theta_a\theta_b}$ be the Kronecker delta function so that,

$$\delta_{\theta_a\theta_b} = \begin{cases} 1 & \text{if } \theta_a=\theta_b=0 \\ 0 & \text{otherwise} \end{cases} \tag{3.8}$$

that is, $\delta_{\theta_a\theta_b}=1$ for the double heterozygotes in coupling or repulsion phase and 0 for any other genotype. This function and variable θ have the following properties:

$$\begin{aligned} &\theta_a\delta_{\theta_a\theta_b}=0 \text{ and } \theta_b\delta_{\theta_a\theta_b}=0; \\ &\theta_a\theta_b\delta_{\theta_a\theta_b}=0. \end{aligned} \tag{3.9}$$

The first property is apparent since $\delta_{\theta_a\theta_b}$ is always zero with the absence of double heterozygotes. $\delta_{\theta_a\theta_b}$ takes values 1 and 0, but $\delta_{\theta_a\theta_b} = 1$ only if $\theta_a = \theta_b = 0$. Thus, $\theta_a\theta_b\delta_{\theta_a\theta_b} = 0$. This proves the second property.

An individual with genotype $\Theta(\theta_a,\theta_b)$ produces gametes containing AB,

Ab, aB and ab in the frequencies displayed in Table 3.1. Here D is linkage disequilibrium (Falconer, 1981, p19) which can be defined as half the difference in frequency between coupling, $2g_1g_4$ and repulsion, $2g_2g_3$, heterozygotes in selfed progenies of $\Theta(\theta_a,\theta_b)$, i.e.,

$$D = g_1g_4 - g_2g_3 , \tag{3.10}$$

and which can be positive or negative, depending on whether the parental population is in coupling or repulsion phase, respectively. For example, if the parental population is the F_1 of a cross between two inbred lines, $AABB \times aabb$, then $g_1 = g_4 = (1 - p)/2$, $g_2 = g_3 = p/2$ and $D = (1-2p)/4$; if the parental population is the F_1 of a cross between two inbred lines $AAbb \times aaBB$ then $g_1 = g_4 = p/2$, $g_2 = g_3 = (1 - p)/2$ and $D = -(1-2p)/4$.

Table 3.1: Gametic frequencies produced by genotype $\Theta(\theta_a.\theta_b)$ with two loci each with two equally frequent alleles

Gamete	Frequency
AB	$g_1 = (\frac{1}{4} + D\delta_{\theta_a\theta_b})(1 + \theta_a + \theta_b + \theta_a\theta_b)$
Ab	$g_2 = (\frac{1}{4} - D\delta_{\theta_a\theta_b})(1 + \theta_a - \theta_b - \theta_a\theta_b)$
aB	$g_3 = (\frac{1}{4} - D\delta_{\theta_a\theta_b})(1 - \theta_a + \theta_b - \theta_a\theta_b)$
ab	$g_4 = (\frac{1}{4} + D\delta_{\theta_a\theta_b})(1 - \theta_a - \theta_b + \theta_a\theta_b)$

Let $\Theta'(\theta_a',\theta_b')$ and $\Theta''(\theta_a'',\theta_b'')$ be the two parental genotypes. Then the cross $\Theta' \times \Theta''$ produces ten genotypes with the frequencies displayed in Table 3.2. Here:

Table 3.2: Genotypic frequencies resulting from a population formed from AA, Ab, aB and ab gametes with frequencies in Table 3.1

Genotype	θ_a	θ_b	Frequency
$AABB$	1	1	$w(1+\theta_a')(1+\theta_a'')(1+\theta_b')(1+\theta_b'')$
$AABb$	1	0	$(1+\theta_a')(1+\theta_a'')[y(1+\theta_b')(1-\theta_b'') + z(1-\theta_b')(1+\theta_b'')]$
$AAbb$	1	−1	$x(1+\theta_a')(1+\theta_a'')(1-\theta_b')(1-\theta_b'')$
$AaBB$	0	1	$(1+\theta_b')(1+\theta_b'')[y(1+\theta_a')(1-\theta_a'') + z(1-\theta_a')(1+\theta_a'')]$
AB/ab	0	0	$w[(1+\theta_a')(1-\theta_a'')(1+\theta_b')(1-\theta_b'') + (1-\theta_a')(1+\theta_a'')(1-\theta_b')(1+\theta_b'')]$
Ab/aB	0	0	$x[(1+\theta_a')(1-\theta_a'')(1-\theta_b')(1+\theta_b'') + (1-\theta_a')(1+\theta_a'')(1+\theta_b')(1-\theta_b'')]$
$Aabb$	0	−1	$(1-\theta_b')(1-\theta_b'')[y(1+\theta_a')(1-\theta_a'') + z(1-\theta_a')(1+\theta_a'')]$
$aaBB$	−1	1	$x(1-\theta_a')(1-\theta_a'')(1+\theta_b')(1+\theta_b'')$
$aaBb$	−1	0	$(1-\theta_a')(1-\theta_a'')[y(1+\theta_b')(1-\theta_b'') + z(1-\theta_b')(1+\theta_b'')]$
$aabb$	−1	−1	$w(1-\theta_a')(1-\theta_a'')(1-\theta_b')(1-\theta_b'')$

$$w = (\frac{1}{4} + D\delta')(\frac{1}{4} + D\delta''),$$
$$x = (\frac{1}{4} - D\delta')(\frac{1}{4} - D\delta''),$$
$$y = (\frac{1}{4} + D\delta')(\frac{1}{4} - D\delta'') \text{ and}$$
$$z = (\frac{1}{4} - D\delta')(\frac{1}{4} + D\delta''),$$

and no confusion will result from omitting subscripts of δ specifying the genotypes in the parental generation.

It is possible to show from Table 3.2 that Genotypes AA, Aa and aa have frequencies $\frac{1}{4}(1+\theta_a')(1+\theta_a'')$, $\frac{1}{2}(1-\theta_a'\theta_a'')$ and $\frac{1}{4}(1-\theta_a')(1-\theta_a'')$, respectively. The expectations of θ and θ^2 in the progeny are

$$E(\theta) = \frac{1}{2}(\theta' + \theta'')$$

$$E(\theta^2) = \frac{1}{2}(1 + \theta'\theta''),$$

as obtained by Hayman (1954).

To this list the expectations of the statistics concerning both loci can be added. The covariances in the progeny, $cov(\theta_a, \theta_b)$, $cov(\theta_a, \theta_b^2)$, $cov(\theta_a^2, \theta_b)$ and $cov(\theta_a^2, \theta_b^2)$, can be derived from Table 3.2. It is now just a matter of algebra to show that

$$\begin{aligned} cov(\theta_a, \theta_b) &= D(\delta' + \delta''), \\ cov(\theta_a, \theta_b^2) &= 0, \\ cov(\theta_a^2, \theta_b) &= 0, \\ cov(\theta_a^2, \theta_b^2) &= 4D^2\delta'\delta''. \end{aligned} \tag{3.11}$$

In the absence of linkage, all four covariances have zero values.

When an individual with genotype $\Theta(\theta_a, \theta_b)$ is selfed, $\theta_a' = \theta_a'' = \theta_a$, $\theta_b' = \theta_b''=\theta_b$ and $\delta' = \delta'' = \delta$. Under continued selfing up to the kth generation, the genotypic frequencies with respect to one locus and two loci can be obtained in usual way.

The one-locus genotypic frequencies after k generations of selfing can be obtained by using an important invariant property of θ given in Hayman (1955),

$$E^k(1-\theta^2) = (\frac{1}{2})^{k-1}E(1-\theta^2),$$

where E^k represents expectation at the kth generation of selfing. The frequencies of genotypes AA, Aa and aa after $k-1$ generations of selfing are as follows,

Genotype	Frequency after $k-1$ generations of selfing
AA	$\frac{1}{2}(1+\theta_a) - (\frac{1}{2})^k(1-\theta_a^2)$
Aa	$(\frac{1}{2})^{k-1}(1-\theta_a^2)$
aa	$\frac{1}{2}(1-\theta_a) - (\frac{1}{2})^k(1-\theta_a^2)$

(3.12)

Now turn to the case of two loci. Let $\boldsymbol{f}^k$ be a vector of two-locus genotypic frequencies after k - 1 generations of selfing from the F_2. It can be written as

$$\begin{aligned} \boldsymbol{f}^k &= \boldsymbol{A}\boldsymbol{f}^{k-1}, \quad \text{for } k \geq 2, \\ &= \boldsymbol{A}^{k-2}\boldsymbol{f}^2, \end{aligned} \tag{3.13}$$

where $\boldsymbol{A}$ is the matrix of transition probabilities which can be defined as the

probabilities of going from one genotype to another in one generation of selfing. For example, the probability of going from genotype $AABB$ to itself is unity while the probability of going from genotype $AABb$ to $AABB$ is only one quarter. Evaluation of $\boldsymbol{A}^{k-2}$ has been given by Nelder (1952).

From Tables 3.1 and 3.2, $\boldsymbol{f}^2$ becomes $\boldsymbol{f}^2 = (g_1^2,\ 2g_1g_2,\ g_2^2,\ 2g_1g_3,\ 2g_1g_4,\ 2g_2g_3,\ 2g_2g_4,\ g_3^2,\ 2g_1g_4,\ g_4^2)'$. Thus, frequencies in F_k generation of selfing, $\boldsymbol{f}^k$, can be obtained by multiplying $\boldsymbol{A}^{k-2}$ from Nelder by $\boldsymbol{f}^2$. Results are displayed in Table 3.3. As a check on this result, the frequency of AA, for example, given in Eqn.(3.12) is simply the sum of the frequencies for $AABB$, $AABb$ and $AAbb$ in Table 3.3.

Note that $\boldsymbol{f}^k$ has been defined with sufficient generality that this definition allows an arbitrary number ($m = 0, 1, 2, \ldots$) of generations of random mating prior to initation of continuous self-fertilization. At the end of these m generations of random mating, linkage disequilibrium $D_m = D_0(1-p)^m$, where D_0 is the initial linkage disequilibrium with zero generations ($m = 0$) of random mating prior to selfing. For gene frequency one half, $D_0 = \pm(1-2p)/4$, where $\pm$ should be read as coupling (+) or repulsion (−) linkages in the F_1, respectively. Furthermore, with this definition, the F_1 may not necessarily be the double heterozygotes and can be completely specified by θ_a, θ_b and δ. For example, for genotype $AABb$, $\theta_a = 1$, $\theta_b = 0$ and $\delta = 0$.

Now consider an individual with genotype $\Theta(\theta_a,\ \theta_b)$. The genotypic value M (in Hayman's notation) can be expressed as a general polynomial function of the genotype, Θ,

Table 3.3: Frequencies of ten genotypes with respect to two linked loci at F_k generation of selfing

Genotype	$f^k = A^{k-2} f^2$
$AABB$	$g_1^2+(g_1g_2+g_1g_3+g_1g_4+g_2g_3)F_k+\frac{g_1g_4-g_2g_3}{2}\sum_{i=1}^{k-2}(\frac{\lambda}{2})^i+\frac{g_1g_4+g_2g_3}{2}[(\frac{\xi}{2})^{k-2}-1]$
$AABb$	$(g_1g_4+g_2g_3+2g_1g_2)(1-F_k)-(g_1g_4+g_2g_3)(\frac{\xi}{2})^{k-2}$
$AAbb$	$g_2^2+(g_1g_2+g_2g_4+g_1g_4+g_2g_3)F_k-\frac{g_1g_4-g_2g_3}{2}\sum_{i=1}^{k-2}(\frac{\lambda}{2})^i+\frac{g_1g_4+g_2g_3}{2}[(\frac{\xi}{2})^{k-2}-1]$
$AaBB$	$(g_1g_4+g_2g_3+2g_1g_3)(1-F_k)-(g_1g_4+g_2g_3)(\frac{\xi}{2})^{k-2}$
AB/ab	$2g_1g_4(\frac{\xi}{2})^{k-2}$
Ab/aB	$2g_2g_3(\frac{\xi}{2})^{k-2}$
$Aabb$	$(g_1g_4+g_2g_3+2g_2g_4)(1-F_k)-(g_1g_4+g_2g_3)(\frac{\xi}{2})^{k-2}$
$aaBB$	$g_3^2+(g_1g_3+g_3g_4+g_1g_4+g_2g_3)F_k-\frac{g_1g_4-g_2g_3}{2}\sum_{i=1}^{k-2}(\frac{\lambda}{2})^i+\frac{g_1g_4+g_2g_3}{2}[(\frac{\xi}{2})^{k-2}-1]$
$aaBb$	$(g_1g_4+g_2g_3+2g_3g_4)(1-F_k)-(g_1g_4+g_2g_3)(\frac{\xi}{2})^{k-2}$
$aabb$	$g_4^2+(g_2g_4+g_3g_4+g_1g_4+g_2g_3)F_k+\frac{g_1g_4-g_2g_3}{2}\sum_{i=1}^{k-2}(\frac{\lambda}{2})^i+\frac{g_1g_4+g_2g_3}{2}[(\frac{\xi}{2})^{k-2}-1]$

Note $F_k = 1-(\frac{1}{2})^{k-2}$

$$M(\theta_a, \theta_b) = c_0 + c_1\theta_a + c_2\theta_b + c_3\theta_a^2 + c_4\theta_b^2 + c_5\theta_a\theta_b$$

$$+ c_6\theta_a\theta_b^2 + c_7\theta_a^2\theta_b + c_8\theta_a^2\theta_b^2, \qquad (3.14)$$

where c's are constants and can be adjusted to provide appropriate models of gene action.

3.1.2.2. Genetic means from selfing

At the end of the last section, a general expression for genotypic value decomposition was given, using Hayman's device. It was remarked that c's in Eqn.(3.14) can be adjusted to provide the various types of gene action, that is to say that the adjusted c's will be components in the genotypic decomposition. Because the definition of these components depends on the genotypic frequencies in a population, these components will be meaningful only if they are defined with reference to that population (the so-called reference population).

The choice of which population serves as the reference population is critical in that if the actual population and the reference population are to be the same then the components in the decomposition will have special properties (for example, orthogonality with the absence of linkage). However, the number of actual populations which one could study is infinitely large so that the possibility that it is the same as the reference population is small. The following quotation from Bulmer (1980, p49-50) serves as an example and as argument for continuous pursuit of the present investigation:

> · · · Suppose for example that two pure lines are crossed and then allowed to self-fertilize in successive generations. The genotype frequencies change in each generation, and it is therefore necessary to choose one of them as the reference population for decomposition, since the parameters in a model must remain constant. It is

> conventional to choose the limiting state towards which the population tends after many generations of selfing as the reference population for the decomposition (the so-called F_∞ metric), but this choice is to some extent arbitrary. In particular we might choose the F_2 generation as the reference population, giving rise to the so-called F_2-metric; · · ·

The argument stemming from the above quotation is that the choice of either the completely inbred generation or the F_2 generation as reference population is arbitrary. This leads to consideration of a generalized reference population.

Let the F_k generation be such a generalized reference population. Inbreeding of this selfed population is measured by Wright's inbreeding coefficient, $F_k = 1 - (1/2)^{k-2}$ for $k \geq 2$. Note that $F_2 = 0$ and $F_\infty = 1$. Thus, this F_k-metric depends, in addition, on inbreeding of the population. Then Eqn.(3.14) can be rewritten in terms of standard notation for gene action (Mather and Jinks, 1971, 1982),

$$\begin{aligned} M(\theta_a, \theta_b; F_k) = m + \theta_a d_a + \theta_b d_b + (\frac{1+F_k}{2} - \theta_a^2)h_a + (\frac{1+F_k}{2} - \theta_b^2)h_b \\ + \theta_a\theta_b i_{ab} + \theta_a(\frac{1+F_k}{2} - \theta_b^2)j_{ab} + (\frac{1+F_k}{2} - \theta_a^2)\theta_b j_{ba} \\ + (\frac{1+F_k}{2} - \theta_a^2)(\frac{1+F_k}{2} - \theta_b^2)l_{ab}\,. \end{aligned} \tag{3.15}$$

Thus, F_2-metric and F_∞-metric can be obtained by substituting 0 and 1 for F_k, respectively. The expectation of $M(\theta_a, \theta_b; F_k)$ is found below

$$E[M(\theta_a, \theta_b; F_k)] = m + E(\theta_a)d_a + E(\theta_b)d_b + E[\frac{1+F_k}{2} - \theta_a^2]h_a$$

$$+ E[\frac{1+F_k}{2} - \theta_b^2]h_b + [E(\theta_a)E(\theta_b) + cov(\theta_a, \theta_b)]i_{ab}$$

$$+ \{E(\theta_a)E[\frac{1+F_k}{2} - \theta_b^2] + cov(\theta_a, \theta_b^2)\}j_{ab} \qquad (3.16)$$

$$+ \{E[\frac{1+F_k}{2} - \theta_a^2]E(\theta_b) + cov(\theta_a^2, \theta_b)\}j_{ba}$$

$$+ \{E[\frac{1+F_k}{2} - \theta_a^2]E[\frac{1+F_k}{2} - \theta_a^2] + cov(\theta_a^2, \theta_b^2)\}l_{ab},$$

where

$$E(\theta_a) = \frac{1}{2}(\theta_a{}' + \theta_a{}''),$$

$$E(\theta_a^2) = \frac{1}{2}(1 + \theta_a{}'\theta_a{}''),$$

$$cov(\theta_a, \theta_b) = D(\delta' + \delta''),$$

$$cov(\theta_a, \theta_b^2) = 0,$$

$$cov(\theta_a^2, \theta_b) = 0 \text{ and}$$

$$cov(\theta_a^2, \theta_b^2) = 4D^2\delta'\delta''.$$

Note for example that

$$E(\theta_a\theta_b) = E(\theta_a)E(\theta_b) + cov(\theta_a, \theta_b). \qquad (3.17)$$

In the absence of linkage, all four *cov*'s have zero values and,

$$
\begin{aligned}
E[M(\theta_a, \theta_b; F_k)] = m &+ \frac{1}{2}(\theta_a' + \theta_a'')d_a + \frac{1}{2}(\theta_b' + \theta_b'')d_b \\
&+ \frac{1}{2}(F_k - \theta_a'\theta_a'')h_a + \frac{1}{2}(F_k - \theta_b'\theta_b'')h_b \\
&+ \frac{1}{4}(\theta_a' + \theta_a'')(\theta_b' + \theta_b'')i_{ab} \\
&+ \frac{1}{4}(\theta_a' + \theta_a'')(F_k - \theta_b'\theta_b'')j_{ab} \\
&+ \frac{1}{4}(F_k - \theta_a'\theta_a'')(\theta_b' + \theta_b'')j_{ba} \\
&+ \frac{1}{4}(F_k - \theta_a'\theta_a'')(F_k - \theta_b'\theta_b'')l_{ab}.
\end{aligned}
\tag{3.18}
$$

Suppose for example that the F_∞-metric is used, that is to say $F_k = 1$. Then each mean of ten genotypes with respect to two linked loci after k - 1 generations of selfing from a single F_1 family can be expressed in the following form:

$$
\begin{aligned}
E^k[M(\theta_a, \theta_b; 1)] = m &+ E^k(\theta_a)d_a + E^k(\theta_b)d_b \\
&+ E^k(1 - \theta_a^2)h_a + E^k(1 - \theta_b^2)h_b \\
&+ [E^k(\theta_a)E^k(\theta_b) + cov^k(\theta_a, \theta_b)]i_{ab} \\
&+ E^k(\theta_a)E^k(1 - \theta_b^2)j_{ab} + E^k(1 - \theta_a^2)]E^k(\theta_b)j_{ba} \\
&+ \{E^k(1 - \theta_a^2)E^k(1 - \theta_a^2) + cov^k(\theta_a^2, \theta_b^2)\}l_{ab},
\end{aligned}
\tag{3.19}
$$

where $E^k[M(\theta_a, \theta_b;1)]$ represents the expected value of genotypic mean obtained from the k - 1 generations of selfing, and $E^k(\theta_a)$ and $cov^k(\theta_a, \theta_b)$, for example, are the expected value of variable θ_a and the covariance between two variables θ_a and θ_b at the k*th* generation of selfing, respectively. From Eqn.(3.12) and Table 3.3,

$$
\begin{aligned}
&E^k(\theta) = \theta\,,\\
&E^k(1-\theta^2) = (\frac{1}{2})^{k-1}(1-\theta^2)\,,\\
&cov^k(\theta_a, \theta_b) = \pm \sum_{i=1}^{k-1} (\frac{\lambda}{2})^i \delta\,,\\
&cov^k(\theta_a^2, \theta_b^2) = [(\frac{\xi}{2})^{k-1} - (\frac{1}{4})^{k-1}]\delta^2\,.
\end{aligned}
\tag{3.20}
$$

Therefore, the explicit expression of each of ten genotypic means after k - 1 generations of selfing is obtained by defining genic effect with respect to the F_∞-metric (i.e., F_k = 1). For example, the mean of progeny of a coupling double heterozygote, AB/ab, at the F_k generation of selfing is

$$E^k[M(0,0;1)] = m + (\frac{1}{2})^{k-1}(h_a + h_b) + (\frac{\lambda}{2-\lambda})[1 - (\frac{\lambda}{2})^{k-1}]i_{ab} + (\frac{\xi}{2})^{k-1} l_{ab}\,.$$

The genotypic value of an individual in an F_k-derived F_n or F_k-derived $F_{n'}$ subpopulation is conditioned by the genotype of the F_k ancestor. This is to say that the value of F_k-derived F_n subpopulation mean is equal to the expected value of F_k-derived F_n subpopulation genotypes, given the genotype of its F_k ancestor. Thus, genotypic means in F_k-derived F_n subpopulations are equal to the expected values of genotypes in F_k after n - k generations of selfing of given F_k ancestral genotypes. The means of all ten genotypes in F_k after these additional n - k generations of selfing with respect to the F_∞-metric are given in Table 3.4.

Table 3.4: F_k-derived F_n subpopulation means for two linked loci

Genotypes	θ_a	θ_b	F_k-derived F_n subpopulation mean
$AABB$	1	1	$d_a + d_b + i_{ab}$
$AABb$	1	0	$d_a + (\frac{1}{2})^{n-k}(h_b + j_{ab})$
$AAbb$	1	−1	$d_a - d_b - i_{ab}$
$AaBB$	0	1	$d_b + (\frac{1}{2})^{n-k}(h_a + j_{ba})$
AB/ab	0	0	$(\frac{1}{2})^{n-k}(h_a + h_b) + \sum_{t=1}^{n-k}(\frac{\lambda}{2})^t i_{ab} + (\frac{\xi}{2})^{n-k} l_{ab}$
Ab/aB	0	0	$(\frac{1}{2})^{n-k}(h_a + h_b) - \sum_{t=1}^{n-k}(\frac{\lambda}{2})^t i_{ab} + (\frac{\xi}{2})^{n-k} l_{ab}$
$Aabb$	0	−1	$-d_b + (\frac{1}{2})^{n-k}(h_a - j_{ba})$
$aaBB$	−1	1	$-d_a + d_b - i_{ab}$
$aaBb$	−1	0	$-d_a + (\frac{1}{2})^{n-k}(h_b - j_{ab})$
$aabb$	−1	−1	$-d_a - d_b + i_{ab}$

3.1.3. Results and discussion

3.1.3.1. Two-locus covariances of relatives from selfing

It is sufficient to derive the expression for $Cov(k,1;n,n')$ under the system of selfing. The other desirable quantities can be obtained by using the relations given in Section 3.1.1. Since Gates *et al.* (1957) and Mather and Jinks (1971, 1982) utilized the F_∞-metric in their derivations, the following derivation will be based on the F_∞-metric.

Following the procedures set forth by Gates *et al.* (1957), but making use of Tables 3.3 and 3.4, a general formula of $Cov(k,1;n,n')$ is obtained for the two-locus model of additive, dominance and epistatic effect with linkage under selfing of generations from a cross between two inbred lines. Eleven distinct quadratic components and the coefficients associated with the components from the formula are displayed in Table 3.5. Three terms in Table 3.5 have a $\pm$ sign which should be read as the coupling (+) or repulsion (−) phase in the F_1 of a cross between two inbred lines. Premultiplying of the 11 quadratic terms by their corresponding coefficents results in the general expression for $Cov(k,1;n,n')$. This expression is subject to the condition $2 \leq k \leq n \leq n'$. In particular, $Cov(2,1;2,2)$ represents the total genetic variance of the F_2 and the expression for $Cov(2,1;2,2)$ becomes that for V_{1F2} of Pooni and Jinks (1982).

One of the assumptions made in Gates' test is that there is no non-allelic interaction (epistasis). It is desirable to show, from the general results given in Table 3.5, if the presence of epistasis has any impact on Gates' test for linkage. Note that the components of genetic variance involving linkage change only when different ranks (simple and/or compound) of covariances have been used. It is then logical to find out if any epistatic variance also changes with rank. The present discussion is limited to the case of simple rank (i.e., $k' = k-1$). For highly inbred progenies from an F_k ancestral generation, let (1) $Cov(1/2;k,k-1;\infty,\infty)$ be the covariances with epistasis, but no linkage; (2) $Cov(p;k,k-1;\infty,\infty)$ be the covariances with coupling linkage, but no epistasis and (3) $Cov(-p;k,k-1;\infty,\infty)$ be the covariances with repulsion linkage, but no epistasis. Then from Table 3.5 and using Eqn.(3.4),

Table 3.5: Quadratic components and their coefficients of two-locus covariance between relatives, $Cov(k,1;n,n')$, from selfing in a cross of two inbred lines

Component	Coefficient
$d_a^2 + d_b^2$	$1 - (\frac{1}{2})^{k-1}$
$d_a d_b$	$\pm 2\,(\frac{\lambda}{2-\lambda})\,[\,1 - (\frac{\lambda}{2})^{k-1}\,]$
$h_a^2 + h_b^2$	$[\,2^{k+1} - 4\,]\,(\frac{1}{2})^{n+n'}$
$h_a h_b$	$[\,2^{k-2}\xi^{k-1} - \frac{1}{2}\,]\,(\frac{1}{2})^{n+n'-4}$
i_{ab}^2	$[\,1 - (\frac{1}{2})^{k-2}\,] + (\frac{\xi}{2})^{k-1}\{\,1 + (\frac{\lambda}{2-\lambda})^2\,[\,1 - (\frac{\lambda}{2})^{n-k}\,]\,[\,1 - (\frac{\lambda}{2})^{n'-k}\,]\,\}$ $- (\frac{\lambda}{2-\lambda})^2\,[\,1 - (\frac{\lambda}{2})^{n-1}\,]\,[\,1 - (\frac{\lambda}{2})^{n'-1}\,]$
$j_{ab}^2 + j_{ba}^2$	$(\,1 - \xi^{k-1}\,)\,(\frac{1}{2})^{n+n'-k-1}$
l_{ab}^2	$(\frac{\xi}{2})^{n+n'-k-1}\,[\,1 - (\frac{\xi}{2})^{k-1}\,]$
$d_a j_{ab} + d_b j_{ba}$	$(\,1 - \xi^{k-1}\,)\,[\,(\frac{1}{2})^{n-1} + (\frac{1}{2})^{n'-1}\,]$
$(h_a + h_b)\,i_{ab}$	$\pm(\frac{\lambda}{2-\lambda})\,\{\,[\,(\frac{1}{2})^{n-k} - (\frac{1}{2})^{n-1}\,]\,[\,(\frac{\lambda}{2})^{k-1} - (\frac{\lambda}{2})^{n-1}\,]$ $+ [\,(\frac{1}{2})^{n'-k} - (\frac{1}{2})^{n'-1}\,]\,[\,(\frac{\lambda}{2})^{k-1} - (\frac{\lambda}{2})^{n'-1}\,]$ $- [\,1 - (\frac{\lambda}{2})^{k-1}\,]\,[\,(\frac{1}{2})^{n-1} + (\frac{1}{2})^{n'-1}\,]\,\}$
$(h_a + h_b)\,l_{ab}$	$[\,1 - (\frac{1}{2})^{k-1}\,]\,[\,(\frac{1}{2})^{n-k}\,(\frac{\xi}{2})^{n'-1} + (\frac{1}{2})^{n'-k}\,(\frac{\xi}{2})^{n-1}\,]$
$i_{ab} l_{ab}$	$\pm(\frac{\lambda}{2-\lambda})\,\{\,[\,(\frac{\xi}{2})^{n'-k} - (\frac{\xi}{2})^{n'-1}\,]\,[\,(\frac{\lambda}{2})^{k-1} - (\frac{\lambda}{2})^{n-1}\,]$ $+ [\,(\frac{\xi}{2})^{n-k} - (\frac{\xi}{2})^{n-1}\,]\,[\,(\frac{\lambda}{2})^{k-1} - (\frac{\lambda}{2})^{n'-1}\,]$ $- [\,1 - (\frac{\lambda}{2})^{k-1}\,]\,[\,(\frac{\xi}{2})^{n-1} + (\frac{\xi}{2})^{n'-1}\,]\,\}$

$$Cov(\frac{1}{2};k,k-1;\infty,\infty) = (\frac{1}{2})^{k-1}(d_a^2 + d_b^2 + (2 - \frac{3}{2^{k-1}})i_{ab}^2)$$
$$= (\frac{1}{2})^{k-1}A$$

$$Cov(p;k,k-1;\infty,\infty) = (\frac{1}{2})^{k-1}[d_a^2 + d_b^2 + 2\lambda^{k-1}d_a d_b]$$
$$= (\frac{1}{2})^{k-1}B \qquad (3.21)$$

$$Cov(-p;k,k-1;\infty,\infty) = (\frac{1}{2})^{k-1}[d_a^2 + d_b^2 - 2\lambda^{k-1}d_a d_b]$$
$$= (\frac{1}{2})^{k-1}C$$

Note that as the rank $(k-1)$ increases A and C increase, but B decreases for $k \geq 2$ and $0 < \lambda < 1$.

Numerical values of A, B and C for λ = 0.0, 0.2, 0.4, 0.6, 0.8 and 1.0 and k = 2 to 10 are given in Table 3.6. The computation takes the form $d_a = d_b = i_{ab} = 1$. This table showed that as k increases contributions of epistasis and repulsion linkage to the covariances increase, while the contribution of coupling linkage decreases. This suggests that the epistatic interaction has similar effects on the covariances as repulsion linkage, thus mimicing the effect of repulsion linkage under any degree of inbreeding. Hayman and Mather (1955) drew the same conclusion based on a specific example.

The negative values of β_3 and β_5 in Gates' test suggest a preponderance of repulsion linkage. For eight traits of a soybean cross, Gates *et al.* (1960) found that repulsion linkages predominanted for height and yield. Because the test is based on the assumption of no epistasis, repulsion linkage for the two traits in their study may also reflect the presence of additive × additive epistatic interaction. By using some of the same genetic

Table 3.6: Comparison between effects of epistasis (A), coupling (B) and repulsion (C) linkages on covariances of selfed relatives

		$\lambda = 1 - 2p$					
k	A†	0.0	0.2	0.4	0.6	0.8	1.0
				B†			
2	2.5000	2.0000	2.4000	2.8000	3.2000	3.6000	4.0000
3	3.2500	2.0000	2.0800	2.3200	2.7200	3.2800	4.0000
4	3.6250	2.0000	2.0160	2.1280	2.4320	3.0240	4.0000
5	3.8125	2.0000	2.0032	2.0512	2.2592	2.8192	4.0000
6	3.9062	2.0000	2.0006	2.0205	2.1555	2.6554	4.0000
7	3.9531	2.0000	2.0001	2.0082	2.0933	2.5243	4.0000
8	3.9766	2.0000	2.0000	2.0033	2.0560	2.4194	4.0000
9	3.9883	2.0000	2.0000	2.0013	2.0336	2.3355	4.0000
10	3.9941	2.0000	2.0000	2.0005	2.0201	2.2684	4.0000
				C†			
2	2.5000	2.0000	1.6000	1.2000	0.8000	0.4000	0.0000
3	3.2500	2.0000	1.9200	1.6800	1.2800	0.7200	0.0000
4	3.6250	2.0000	1.9840	1.8720	1.5680	0.9760	0.0000
5	3.8125	2.0000	1.9968	1.9488	1.7408	1.1808	0.0000
6	3.9062	2.0000	1.9994	1.9795	1.8445	1.3446	0.0000
7	3.9531	2.0000	1.9999	1.9918	1.9067	1.4757	0.0000
8	3.9766	2.0000	2.0000	1.9967	1.9440	1.5806	0.0000
9	3.9883	2.0000	2.0000	1.9987	1.9664	1.6645	0.0000
10	3.9941	2.0000	2.0000	1.9995	1.9799	1.7316	0.0000

† A, B and C are defined in Eqn.(3.21)

material of Gates *et al.* (1960), but using a different model (additive and additive × additive), Hanson and Weber (1961) found significant additive × additive variance for seed weight and percent oil. Hanson and Weber then concluded that epistatic variability could be an important source of genetic variability for seed yield.

If there is no linkage ($\lambda = 0$), then four terms $d_a d_b$, $h_a h_b$, $(h_a + h_b) i_{ab}$ and $i_{ab} l_{ab}$ in Table 3.5 will be zero since their coefficients are functions of linkage only. In this case, just seven terms remain. This feature was also noted by Wright (1987, 1988) who used the identity by descent theory to develop the expression for covariance of relatives from selfing under assumptions of linkage and identity equilibria. On the other hand, if there is a complete linkage ($\lambda = 1$), the two terms in Table 3.5, $(j_{ab}^2 + j_{ba}^2)$ and $(d_a j_{ab} + d_b j_{ba})$, will have zero values.

Rather than dealing with the case of multiple alleles, the present study is limited to the case of two equally frequent alleles per locus. This single assumption yields a considerable simplification in deriving the general expression for $Cov(k,1;n,n')$. This assumption is also made in Gates' test for linkage (Gates *et al.*, 1960).

The more general approach to this derivation is suggested by Weir and Cockerham (1977) who gave the unwieldy results of two-locus variance for the case of multiple alleles. Because Weir and Cockerham used the two-locus identity by descent (*ibd*) theory in their derivation of the expression, evaluation of the coefficients associated with the quadratic components becomes much more complicated, although this may be necessary for the case of multiple alleles. In the presence of linkage, evaluation of the coefficients

of the quadratic components in covariances between inbred relatives by use of *ibd* measures has been limited to the case of additive × additive variance with linkage equilibrium (Cockerham, 1984b; Cockerham and Matzinger, 1985) and has been ignored by Wright (1987, 1988). In terms of experimental estimation, these results are of little value.

Furthermore, the assumption of linkage equilibrium in the initial population of self-fertilized species (Cockerham, 1983; Wright, 1987) is questionable. For an F_2 population derived from a cross between two inbred lines, linkage disequilibrium must surely exist. Thus, Cockerham's model may be useful only for dealing with inbreds derived (by selfing) from a cross-fertilized base population where the assumption of linkage equilibrium can be relaxed.

3.1.3.2. Generalized Gates' test for linkage

When there is no epistasis (i.e. $i = j = l = 0$), only the first four components in the covariance of relatives displayed in Table 3.5 have nonzero values. Thus,

$$\begin{aligned} Cov(k,1;n,n') = {} & [1-(\tfrac{1}{2})^{k-1}](d_a^2+d_b^2) \\ & + 2(\frac{1-2p}{1+2p})[1-(\frac{1-2p}{2})^{k-1}]d_a d_b \\ & + [2^{k+1}-4](\tfrac{1}{2})^{k+n'}(h_a^2+h_b^2) \\ & + [2^{k-2}(1-2pq)^{k-1}-\tfrac{1}{2}](\tfrac{1}{2})^{m+n'-4}h_a h_b . \end{aligned} \tag{3.22}$$

and $Cov(k',1;n,n')$ can be obtained by merely substituting k' for k in Eqn.(3.22). Hence,

$$
\begin{aligned}
Cov(k,k';n,n') &= Cov(k,1;n,n') - Cov(k',1;n,n') \\
&= [(\tfrac{1}{2})^{k'-1} - (\tfrac{1}{2})^{k-1}](d_a^2 + d_b^2) \\
&\quad + 2(\frac{1-2p}{1+2p})[(\frac{1-2p}{2})^{k'-1} - (\frac{1-2p}{2})^{k-1}]d_a d_b \\
&\quad + (\tfrac{1}{2})^{n+n'-1}[2^k - 2^{k'}](h_a^2 + h_b^2) \\
&\quad + (\tfrac{1}{2})^{n+n'-4}[2^{k-2}(1-2pq)^{k-1} \\
&\quad - 2^{k'-2}(1-2pq)^{k'-1}]h_a h_b ,
\end{aligned}
\tag{3.23}
$$

where $1 \leq k' < k \leq n \leq n'$.

In terms of Gates' test for linkage,

$$Cov(k,k';n,n') \simeq \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 ,$$

where β_1 to β_6 are various components of gene action and linkage which are given explicitly in Eqn.(2.10) , and X_j are defined as follows:

$$X_1 = \sum_{i=k'+1}^{k} (\frac{1}{2})^{i-2} = (\frac{1}{2})^{k'-2} - (\frac{1}{2})^{k-2}$$

$$X_2 = \frac{\sum_{i=k'+1}^{k} 2^{i-2}}{2^{n+n'-4}} = \frac{2^{k-1} - 2^{k'-1}}{2^{n+n'-4}}$$

$$X_3 = -\sum_{i=k'+1}^{k} \binom{i-2}{1} (\frac{1}{2})^{i-3} = \frac{k}{2^{k-3}} - \frac{k'}{2^{k'-3}}$$

$$X_4 = \frac{-\sum_{i=k'+1}^{k} \binom{i-2}{1} 2^{i-2}}{2^{n+n'-4}} = \frac{-2^{k-1}(k-3) + 2^{k'-1}(k'-3)}{2^{n+n'-4}}$$

$$X_5 = \sum_{i=k'+1}^{k} \binom{i-2}{2} (\frac{1}{2})^{i-4} = \frac{k'^2 - k' + 2}{2^{k'-3}} - \frac{k^2 - k + 2}{2^{k-3}}$$

$$X_6 = \frac{\sum_{i=k'+1}^{k} \binom{i-2}{2} 2^{i-2}}{2^{n+n'-4}} = \frac{2^{k-2}(k^2 - 7k + 14)}{2^{n+n'-4}} - \frac{2^{k'-2}(k'^2 - 7k' + 14)}{2^{n+n'-4}}$$

Note that the six coefficients are simply functions of k, k', n and n'. When $k' = 1$ the X_j are the same as given by Gates *et al.* (1960).

As an illustration of the utility of this more general procedure, the 30 covariances were taken from the Table 2 of Gates *et al.* (1960) and the six coefficients for each covariance were calculated. Among these 30 covariances, k' is either 1 or $k-1$. For example, $Cov(3,1;4,4) - Cov(2,1;4,4) = Cov(3,2;4,4)$, thus $k' = 2$ for this particular covariance. The results of these calculations (Table 3.7) are identical to the left side of the Table 2 of Gates *et al.* (1960), as expected.

Table 3.7: Thirty covariances and their coefficients from expectations used in Gates' test for linkage

	Coefficient from expectation					
$Cov(k,k';n,n')$	X_1	X_2	X_3	X_4	X_5	X_6
Cov(2,1;2,3)	1.0000	0.5000	0.0000	0.0000	0.0000	0.0000
Cov(2,1;2,4)	1.0000	0.2500	0.0000	0.0000	0.0000	0.0000
Cov(2,1;2,5)	1.0000	0.1250	0.0000	0.0000	0.0000	0.0000
Cov(2,1;2,6)	1.0000	0.0625	0.0000	0.0000	0.0000	0.0000
Cov(2,1;2,7)	1.0000	0.0313	0.0000	0.0000	0.0000	0.0000
Cov(2,1;3,4)	1.0000	0.1250	0.0000	0.0000	0.0000	0.0000
Cov(2,1;3,5)	1.0000	0.0625	0.0000	0.0000	0.0000	0.0000
Cov(2,1;3,6)	1.0000	0.0313	0.0000	0.0000	0.0000	0.0000
Cov(2,1;3,7)	1.0000	0.0156	0.0000	0.0000	0.0000	0.0000
Cov(3,1;4,5)	1.5000	0.0938	-1.0000	-0.0625	0.0000	0.0000
Cov(3,1;4,6)	1.5000	0.0469	-1.0000	-0.0313	0.0000	0.0000
Cov(3,1;4,7)	1.5000	0.0234	-1.0000	-0.0156	0.0000	0.0000
Cov(4,1;5,6)	1.7500	0.0547	-2.0000	-0.0781	1.0000	0.0313
Cov(4,1;5,7)	1.7500	0.0273	-2.0000	-0.0391	1.0000	0.0156
Cov(5,1;6,7)	1.8750	0.0293	-2.7500	-0.0664	2.5000	0.0547
Cov(2,1;4,5)	1.0000	0.0313	0.0000	0.0000	0.0000	0.0000
Cov(2,1;4,6)	1.0000	0.0156	0.0000	0.0000	0.0000	0.0000
Cov(2,1;4,7)	1.0000	0.0078	0.0000	0.0000	0.0000	0.0000
Cov(3,1;5,6)	1.5000	0.0234	-1.0000	-0.0156	0.0000	0.0000
Cov(3,1;5,7)	1.5000	0.0117	-1.0000	-0.0078	0.0000	0.0000
Cov(4,1;6,7)	1.7500	0.0137	-2.0000	-0.0195	1.0000	0.0078
Cov(2,1;3,3)	1.0000	0.2500	0.0000	0.0000	0.0000	0.0000
Cov(2,1;4,4)	1.0000	0.0625	0.0000	0.0000	0.0000	0.0000
Cov(3,2;4,4)	0.5000	0.1250	-1.0000	-0.1250	0.0000	0.0000
Cov(3,1;5,5)	1.5000	0.0469	-1.0000	-0.0313	0.0000	0.0000
Cov(4,3;5,5)	0.2500	0.0625	-1.0000	-0.1250	1.0000	0.0625
Cov(4,1;6,6)	1.7500	0.0273	-2.0000	-0.0391	1.0000	0.0156
Cov(5,1;7,7)	1.8750	0.0146	-2.7500	-0.0332	2.5000	0.0273
Cov(5,4;6,6)	0.1250	0.0313	-0.7500	-0.0938	1.5000	0.0938
Cov(6,5;7,7)	0.0625	0.0156	-0.5000	-0.0625	1.5000	0.0938

In order to detect linkage from the analysis of covariance of relatives by using Gates' test, one important part of the test is to calculate the six coefficients for the covariances. This is required for the subsequent multiple regression. The procedure of Gates' test allows direct evaluation of $Cov(k,1;n,n')$, but only indirect evaluation of $Cov(k,k';n,n')$. The procedure outlined above overcomes this deficiency by allowing direct evaluation of $Cov(k,k';n,n')$ for any k' values where $1 \leq k' < k$. This more general procedure is particularly useful for simulating an evaluation of Gates' test (See Section 3.2).

3.1.3.3. Bias due to approximation of Gates' test

Gates' test results from an approximate expansion of $Cov(k,1;n,n')$, where $2 \leq k \leq n \leq n'$. Since only the first three terms were taken from the exact expansion of the coefficients for cross products of additive and dominance effects in $Cov(k,1;n,n')$, X_1 to X_6 do not represent the exact expectations of $Cov(k,1;n,n')$ for $k \geq 5$. The deviation of the approximation from the exact expectation should give some indication of the adequacy of Gates' test.

Gates *et al.* (1960) defined that, for a pair of loci,

$$A_1 = \frac{1}{2}(d_a^2 + d_b^2),$$

$$D_1 = \frac{1}{4}(h_a^2 + h_b^2)$$

$$A_{12} = (1-2p)d_a d_b,$$

$$D_{12} = 2(\frac{1-2p}{2})^2 h_a h_b$$

Thus, Eqn.(3.22) leads to

$$Cov(k,1;n,n') = A_1Z_1 + A_{12}Z_2 + D_1Z_3 + D_{12}Z_4,$$

where

$$Z_1 = 2 - \frac{1}{2^{k-2}}$$

$$Z_2 = \sum_{i=2}^{k} \left(\frac{1-2p}{2}\right)^{i-2}$$

$$Z_3 = \frac{2^{k-1}}{2^{n+n'-4}}$$

$$Z_4 = \sum_{i=2}^{k} \frac{2^{i-2}(1-2pq)^{i-2}}{2^{n+n'-4}}.$$

Gates *et al.* (1960) used $X_1 + X_3p + X_5p^2$ to approximate Z_2 and $X_2 + 2X_4pq + 4X_6p^2q^2$ to approximate Z_4. If two ratios are defined as follows,

$$R_1 = \frac{A_{12}(X_1 + X_3p + X_5p^2)}{A_{12}Z_2} = \frac{X_1 + X_3p + X_5p^2}{Z_2}$$

$$R_2 = \frac{D_{12}(X_2 + 2X_4pq + 4X_6p^2q^2)}{D_{12}Z_4} = \frac{X_2 + 2X_4pq + 4X_6p^2q^2}{Z_4}, \qquad (3.24)$$

then R_1 and R_2 may represent the measures of deviations of the approximations from the exact expectations for terms involving linkage in Gates' test. Both ratios have the expectations of unity only for $k < 5$ and $\lambda = 1 - 2p = 1$, as pointed out by Gates *et al.* (1960).

Numerical values of the two ratios for $k = 2$ to 10 generations, and $\lambda = 0.0$, 0.2, 0.4, 0.6, 0.8 and 1.0 are given in Table 3.8. As expected, the ratios for $\lambda = 1.0$ equal unity and those for $k = 2$, 3 and 4 also equal unity at all λ values. All values for the ratios are positive. This is not surprising

Table 3.8: Ratios of approximate to exact expectations for both additive (R_1) and dominance (R_2) cross products in Gates' test

	$\lambda = 1 - 2p$					
k	0.0	0.2	0.4	0.6	0.8	1.0
			R_1 (Additive)			
2	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
3	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
4	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
5	1.125000	1.057606	1.021635	1.005646	1.000616	1.000000
6	1.312500	1.149761	1.058339	1.015718	1.001758	1.000000
7	1.500000	1.245376	1.098070	1.027124	1.003109	1.000000
8	1.656250	1.326938	1.133015	1.037539	1.004396	1.000000
9	1.773438	1.389094	1.160253	1.045912	1.005471	1.000000
10	1.855469	1.433109	1.179876	1.052099	1.006295	1.000000
			R_2 (dominance)			
2	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
3	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
4	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
5	1.250000	1.208347	1.116985	1.038980	1.004790	1.000000
6	2.600000	2.313304	1.703267	1.215997	1.023748	1.000000
7	7.500000	6.250227	3.671588	1.750536	1.073078	1.000000
8	22.428571	18.019283	9.201574	3.092352	1.178799	1.000000
9	63.625000	49.874703	23.235017	6.116750	1.380754	1.000000
10	170.333333	130.780570	56.576964	12.465408	1.738011	1.000000

since X_1 to X_6 were derived from the first three terms of exact binomial expansion where the first and the third terms are positive and the second is negative. The ratios increase as k is increased, but the magnitude of the increase is much greater for R_2 than for R_1, particularly when k is high (> 7). The higher values for the ratios indicate the upward bias caused by the approximate expansion in Gates' test. Although Gates *et al.* (1960) argued that the deviation due to the approximation should be trivial for D_{12} when $k \geq 5$, (only 1/512 for $k = 5$) because of even higher values of n and n', such deviation in A_{12} may be substantial. This is particularly true for $p = 0.5$, i.e. genes at the two loci are segregating independently. For example, when $k = 5$, the deviation is 12.5%, and when $k = 10$, the deviation is about 86%!

3.2. Computer simulation

3.2.1. Materials and methods

3.2.1.1. Description of genetic models

Twenty models with differing gene action were specified for the purpose of simulation (Table 3.9). The first ten involve no epistasis. Models 11 to 19 were those with classical epistatic ratios (Strickberger, 1968). Note that Model 12 is a modification of Model 11 with $0 \leq w \leq 1$, which is called the partial complementary model (Mulitze, 1983). Model 20 is hypothetical with gene action specified by $d_a = d_b = h_a = h_b = i = j = l$. x in Table 3.9 represents any non-zero integer value. The choice of x is arbitrary. It is conventional to take x = 1.

The subdivision of genotypic variance into components due to various

Table 3.9: Coded genotypic values and F_2 ratios with respect to two independently segregating loci for the twenty genetic models (M#)

	Genotype									
M#	*AABB*	*AABb*	*AAbb*	*AaBB*	*AaBb*	*Aabb*	*aaBB*	*aaBb*	*aabb*	F_2 independent genetic ratio
1	4x	3x	2x	3x	2x	x	2x	x	0	1 : 4 : 6 : 4 : 1
2	4x	4x	2x	4x	4x	2x	2x	2x	0	9 : 6 : 1
3	4x	2x	2x	2x	0	0	2x	0	0	1 : 6 : 9
4	4x	5x	2x	5x	6x	3x	2x	3x	0	4 : 4 : 1 : 4 : 2 : 1
5	4x	x	2x	x	-2x	-x	2x	2x	0	1 : 2 : 4 : 1 : 4 : 4
6	6x	4x	2x	5x	3x	x	4x	2x	0	1 : 2 : 3 : 4 : 3 : 2 : 1
7	6x	6x	2x	6x	6x	2x	4x	4x	0	9 : 3 : 3 : 1
8	6x	2x	2x	4x	0	0	4x	0	0	1 : 3 : 3 : 9
9	6x	8x	2x	7x	9x	3x	4x	6x	0	4 : 2 : 2 : 3 : 1 : 2 : 1 : 1
10	6x	0	2x	3x	-3x	-x	4x	-2x	0	1 : 1 : 2 : 1 : 3 : 2 : 2 : 4
11	x	x	0	x	x	0	0	0	0	9 : 7
12	x	wx	0	wx	wx	0	0	0	0	1 : 8 : 7
13	x	x	x	x	x	x	x	x	0	15 : 1
14	x	x	0	x	x	0	x	x	x	13 : 3
15	2x	2x	0	2x	2x	0	x	x	x	9 : 3 : 4
16	2x	2x	2x	2x	2x	2x	x	x	0	12 : 3 : 1
17	2x	x	0	2x	x	0	2x	2x	2x	7 : 6 : 3
18	2x	3x	0	2x	3x	0	x	x	0	6 : 3 : 3 : 4
19	x	3x	2x	x	3x	2x	3x	0	2x	7 : 4 : 3 : 2
20	2x	2x	0	2x	2x	0	0	0	x	9 : 1 : 6

types of gene action can be achieved easily by the use of F_2-metric or F_∞-metric. In order to assess the relative importance of components of total genetic variance in the F_2 population, the F_2-metric will be used so that the components will be mutually orthogonal. With $x = 1$ and $w = 0.5$ (for model 12), the 20 models can be translated in terms of relation between gene effects (Table 3.10). It should be noted that the relations between genic effects given in Table 3.10 are different from those of Mather and Jinks (1982) who used the F_∞-metric in deriving the relations.

Since the genetic variance in F_2 population can be expressed as a linear function of the squared genic effects, the additive effects appearing in each model can be adjusted such that the total genetic variance of the F_2 population, $Cov(2,1;2,2)$, for each model will be unity. The results of these adjustments are given in the last column of Table 3.10. With the adjusted additive effects and the relations specified in Table 3.10, eight components of total genetic variance in the F_2 population were calculated for each of 20 models. The results of these calculations are given in Table 3.11. Note that with a gene frequency of one-half, $\sigma^2_{Aa} = d^2_a/2$, $\sigma^2_{Ab} = d^2_b/2$, $\sigma^2_{Da} = h^2_a/4$, $\sigma^2_{Db} = h^2_b/4$, $\sigma^2_{AA} = i^2_{ab}/4$, $\sigma^2_{AD} = j^2_{ab}/8$, $\sigma^2_{DA} = j^2_{ba}/8$ and $\sigma^2_{DD} = l^2_{ab}/16$. Since each row of this Table sums up to unity, the value of each entry represents the proportion of total variance due to a particular component. Models 11 to 20 show various degrees of epistasis in terms of percentage of contribution of epistatic variance to the total variance. The highest ($> 60\%$) are models 13 and 19, although the constitution of the four epistatic components in the two models is quite different (Table 3.11).

Table 3.10: Relationships among eight genic effects derived from the two locus F_2-metric with the adjusted additive effect of locus A for $Cov(2,1;2,2)$ equal to unity

M#†	Relationship	Adjusted
1	$d_a=d_b;\ h_a=h_b=i=j=l=0$	$d_a = 1.0000$
2	$d_a=d_b=h_a=h_b;\ i=j=l=0$	$d_a = 0.8165$
3	$d_a=d_b=-h_a=-h_b;\ i=j=l=0$	$d_a = 0.8165$
4	$d_a=d_b=1/2h_a=1/2h_b;\ i=j=l=0$	$d_a = 0.5774$
5	$d_a=d_b=-1/2h_a=-1/2h_b;\ i=j=l=0$	$d_a = 0.5774$
6	$d_a=1/2d_b;\ h_a=h_b=i=j=l=0$	$d_a = 0.6325$
7	$d_a=1/2d_b=h_a=1/2h_b;\ i=j=l=0$	$d_a = 0.5164$
8	$d_a=1/2d_b=-h_a=-1/2h_b;\ i=j=l=0$	$d_a = 0.5164$
9	$d_a=1/2d_b=1/2h_a=1/4h_b;\ i=j=l=0$	$d_a = 0.3651$
10	$d_a=1/2d_b=-1/2h_a=-1/4h_b;\ i=j=l=0$	$d_a = 0.3651$
11	$d_a=d_b=h_a=h_b=3/2(i=j=l)$	$d_a = 0.7559$
12	$d_a=d_b=2h_a=2h_b=i=l;\ j=0$	$d_a = 0.8341$
13	$d_a=d_b=h_a=h_b=-1/2(i=j=l)$	$d_a = 0.5164$
14	$-d_a=1/3d_b=-h_a=1/3h_b=1/2(i=j=l)$	$d_a = 0.3203$
15	$d_a=1/3d_b=h_a=1/3h_b=1/2(i=j=l)$	$d_a = 0.3203$
16	$d_a=5d_b=h_a=5h_b=-5/2(i=j=l)$	$d_a = 1.0721$
17	$-d_a=2/3d_b=-h_a=i=j_{ba};\ h_b=j_{ab}=l=0$	$d_a = 0.6667$
18	$d_a=5/7d_b=h_a=5/13h_b=5/2(i=j_{ba})=5/6(j_{ab}=l)$	$d_a = 0.5164$
19	$d_a=-2d_b=h_a=h_b=-i=-j_{ba}=1/4(j_{ab}=l)$	$d_a = 0.4714$
20	$d_a=d_b=h_a=h_b=i=j=l$	$d_a = 0.6963$

† Genetic models from Table 3.9.

Table 3.11: Eight components of genetic variance of F_2 population derived from the two locus F_2-metric for 20 models(M#)

	Locus A		Locus B		Interaction			
M#†	σ^2_{Aa}	σ^2_{Da}	σ^2_{Ab}	σ^2_{Db}	σ^2_{AA}	σ^2_{AD}	σ^2_{DA}	σ^2_{DD}
1	0.500	0.000	0.500	0.000	0.000	0.000	0.000	0.000
2	0.333	0.167	0.333	0.167	0.000	0.000	0.000	0.000
3	0.333	0.167	0.333	0.167	0.000	0.000	0.000	0.000
4	0.167	0.333	0.167	0.333	0.000	0.000	0.000	0.000
5	0.167	0.333	0.167	0.333	0.000	0.000	0.000	0.000
6	0.200	0.000	0.800	0.000	0.000	0.000	0.000	0.000
7	0.133	0.067	0.533	0.267	0.000	0.000	0.000	0.000
8	0.133	0.067	0.533	0.267	0.000	0.000	0.000	0.000
9	0.067	0.133	0.267	0.533	0.000	0.000	0.000	0.000
10	0.067	0.133	0.267	0.533	0.000	0.000	0.000	0.000
11	0.286	0.143	0.286	0.143	0.064	0.032	0.032	0.016
12	0.348	0.043	0.348	0.043	0.174	0.000	0.000	0.043
13	0.133	0.067	0.133	0.067	0.267	0.133	0.133	0.067
14	0.051	0.026	0.462	0.231	0.103	0.051	0.051	0.026
15	0.051	0.026	0.462	0.231	0.103	0.051	0.051	0.026
16	0.575	0.287	0.023	0.012	0.046	0.023	0.023	0.012
17	0.222	0.111	0.500	0.000	0.111	0.000	0.056	0.000
18	0.133	0.067	0.261	0.451	0.011	0.048	0.005	0.024
19	0.111	0.056	0.028	0.056	0.056	0.444	0.028	0.222
20	0.242	0.121	0.242	0.121	0.121	0.061	0.061	0.030

† Genetic models from Table 3.9.

3.2.1.2. Program details

To provide numerical evaluation of Gates' test for linkage, a Turbo PASCAL (1986) program, GATESTEST.PAS, was written for an AT&T PC 6300 microcomputer with three major procedures (See Appendix A for program listing).

Procedure COVS_PAS is capable of simulating genetic variances and covariance with both linkage and epistasis for a pair of loci each with two equally frequent alleles. Input parameters for COVS_PAS included initial linkage phases (coupling or repulsion), eight linkage values $\lambda = (1 - 2p) = 0.0$, 0.2, 0.4, 0.6, 0.8, 0.9, 0.95 and 0.99, 20 genetic models in terms of their genic effects, 30 sets of generation numbers for $Cov(k,k';n,n')$ which were those of Gates *et al.* (1960). Thus, the total numbers of $Cov(k,k';n,n')$ generated by COVS_PAS were 9000($15 \times 20 \times 30$).

A debugging test for COVS_PAS was performed by use of a program, COVB.PAS, provided by Dr. R.J. Baker. Since the simulation is deterministic, COVS_PAS was also checked through a few hand-calculated examples.

Procedure GATES_PAS is to apply Gates' test to analyze the 9000 covariances under each set of combinations of generations (inbreeding), linkage and epistasis. Input parameters were 300 sets of 30 covariances read from a file, a set of 30 expected values of six coefficients, and number of parameters in the reduced model. In the present context, the null hypothesis is that linkage is absent under a model of additive and dominance effects. Thus, the number of parameters in the reduced model is two. The coefficient of determination (R^2) was calculated as proportion of variability among 30

covariances explained by additive and dominance effects by GATES_PAS under each genetic situation. It should be noted that Gates *et al.* (1960) used the uncorrected total sum of squares for determining the significance of the terms associated with linkage as well as additive and dominance effects. GATES_PAS followed the same procedure in calculating R^2. GATES_PAS was validated by the multiple regression analysis procedures in both SAS and MINITAB which are coded in the mainframe computer system (VAX 8650) at the University of Saskatchewan.

Procedure SAMPLING_PAS is to add random components to each chosen set of simulated $Cov(k,k';n,n')$ and then estimate the probability of rejecting the null hypothesis that there is no linkage. To simulate the random errors, it is assumed that the errors which are added to the simulated (model) $Cov(k,k';n,n')$ are normally and independently distributed with mean zero and variance σ^2_{Cov}, where σ^2_{Cov} is assumed to be a function of the model $Cov(k,k';n,n')$, and can be expressed as $\sigma_{Cov} = c \times Cov(k,k';n,n')$, where c is a constant. In order to detect linkage effect, the c value must be much less than 0.5 (see Section 3.2.1.3 for full discussion). This procedure was designed to determine how small c must be to assure detection of linkage in the presence of sampling variability of covariance estimates. With the model $Cov(k,k';n,n')$ and c value, SAMPLING_PAS determined the phenotypic value of $Cov(k,k';n,n')$ as: $P = Cov(k,k';n,n') + N(0, c \times Cov(k,k';n,n'))$.

The probability of rejecting the null hypothesis that there is no linkage under each given genetic situation was estimated through replicated computer runs. The rejection for each run was made according to the usual F criterion

from the analysis of variance in the multiple regression analysis. The number of replicates was determined according to the relation $N > (1.0/\Delta p)^2$, where Δp is the absolute difference between the estimated and the true probabilities. The rationale of this relation is presented in Appendix B. In the present case, $N = 1000$ was chosen such that $\Delta p = 0.03$ (3%). Models #2 (9:6:1 F_2 ratio) and #13 (15:1 F_2 ratio) with linkage values $\lambda =$ 0.0, 0.2, 0.6 and 0.9 for both coupling and repulsion phases were simulated. The c values were chosen as 0.25, 0.125, 0.0625 and 0.03125.

3.2.1.3. Precision factor for detecting linkage

In the previous section, the constant c was introduced as ratio of the standard error of a simulated covariance to the covariance itself for Gates' test. This simulated $Cov(k,k';n,n')$ can be expressed as a linear function of variance components due to additive, dominance, epistatic and linkage effects. To detect these genetic effects through the analysis of a set of covariances, c must be chosen to ensure that the precision of estimated (or simulated) covariances is sufficiently high for such detection. Thus, c may be called the *precision factor.* To detect linkage through the analysis of a set of covariances, the choice of c depends on the relative contribution of linkage effects to the covariances. Suppose for example that linkage is the only component of the covariances. A common practice is that if estimates of such covariances are greater than two times their standard errors, then it is likely that linkage is present. By definition, c, in this case, should be less 0.5. However, since additive and dominance are likely more important components of the covariances than linkage effect, c must be much less than 0.5 to ensure detection of linkage.

3.2.2. Results and discussion

Gates' test is a test of the null hypothesis of no linkage and utilizes the analysis of variance in the multiple regression analysis. Rejection of the hypothesis suggests the presence of linkage under the assumption of no epistasis. Analysis of 300 sets of $Cov(k,k';n,n')$, generated by COVS_PAS, provides for each set of 30 $Cov(k,k';n,n')$ the proportion, R^2, of total genetic variability among these covariances explained only by additive and dominance effects. These R^2 values are reported for eight values of coupling linkage in Table 3.12, and eight values of repulsion linkage in Table 3.13.

As expected, Models (1 to 10) [additive and/or dominance models with no linkage ($\lambda = 0$)] have R^2 values of unity. The models with epistasis (M11 to M20) but no linkage show reductions in R^2. Model 19 (7:4:3:2 F_2 ratio) shows severe reduction in R^2 (about 14%). Model 13 (15:1 F_2 ratio) reduces R^2 by about 2%. The remaining eight models reduces R^2 by less than 1%. With a particular model (e.g. M19 and M13), epistasis can mimic linkage effect. If epistasis is assumed absent, then linkage effect has its maximum impact on R^2 at moderate to tight linkage values ($\lambda = 0.4$ to 0.8). As linkage intensity increases toward $\lambda = 0.99$, the linkage effect seems to disappear. This is true for both coupling and repulsion linkages with the exception of Model #1 with repulsion linkage. In this model, the pattern seems straightforward, but is difficult to interpret. As linkage values increase, the R^2 value decreases. Apart from this exception, all other results can be explained by the apparent fact that with no linkage, two loci act independently, while with complete linkage two loci act as one locus; in either case, linkage cannot be detected.

Table 3.12: Proportion of total variability among 30 covariances, $Cov(k,k';n,n')$, explained by additive and dominance effects for 20 models(M#) with different values of coupling linkage (λ)

	$\lambda = 1 - 2p$							
M#†	0.00	0.20	0.40	0.60	0.80	0.90	0.95	0.99
1	1.0000	0.9993	0.9986	0.9988	0.9995	0.9999	1.0000	1.0000
2	1.0000	0.9993	0.9986	0.9988	0.9995	0.9998	1.0000	1.0000
3	1.0000	0.9993	0.9986	0.9988	0.9995	0.9998	1.0000	1.0000
4	1.0000	0.9994	0.9987	0.9988	0.9995	0.9998	1.0000	1.0000
5	1.0000	0.9994	0.9987	0.9988	0.9995	0.9998	1.0000	1.0000
6	1.0000	0.9995	0.9990	0.9991	0.9996	0.9999	1.0000	1.0000
7	1.0000	0.9995	0.9990	0.9991	0.9996	0.9999	1.0000	1.0000
8	1.0000	0.9995	0.9990	0.9991	0.9996	0.9999	1.0000	1.0000
9	1.0000	0.9996	0.9991	0.9991	0.9996	0.9999	1.0000	1.0000
10	1.0000	0.9996	0.9991	0.9991	0.9996	0.9999	1.0000	1.0000
11	0.9985	0.9992	0.9996	0.9999	1.0000	1.0000	1.0000	1.0000
12	0.9965	0.9989	0.9998	1.0000	1.0000	1.0000	1.0000	1.0000
13	0.9797	0.9921	0.9977	0.9996	1.0000	1.0000	1.0000	1.0000
14	0.9968	0.9922	0.9864	0.9798	0.9724	0.9684	0.9663	0.9645
15	0.9976	0.9971	0.9966	0.9965	0.9975	0.9988	0.9995	1.0000
16	0.9994	0.9996	0.9997	0.9997	0.9998	0.9999	1.0000	1.0000
17	0.9989	0.9973	0.9950	0.9920	0.9886	0.9883	0.9910	0.9986
18	0.9984	0.9987	0.9991	0.9995	0.9998	1.0000	1.0000	1.0000
19	0.8623	0.8602	0.8719	0.9013	0.9536	0.9839	0.9953	0.9998
20	0.9957	0.9964	0.9969	0.9974	0.9984	0.9993	0.9997	1.0000

† Genetic models from Table 3.9.

Table 3.13: Proportion of total variability among 30 covariances, $Cov(k,k';n,n')$, explained by additive and dominance effects for 20 models(M#) with different values of repulsion linkage (λ)

	$\lambda = 1 - 2p$							
M#†	0.00	0.20	0.40	0.60	0.80	0.90	0.95	0.99
1	1.0000	0.9987	0.9946	0.9877	0.9776	0.9714	0.9680	0.9649
2	1.0000	0.9988	0.9952	0.9896	0.9845	0.9867	0.9927	0.9994
3	1.0000	0.9988	0.9952	0.9896	0.9845	0.9867	0.9927	0.9994
4	1.0000	0.9990	0.9966	0.9941	0.9951	0.9979	0.9993	1.0000
5	1.0000	0.9990	0.9966	0.9941	0.9951	0.9979	0.9993	1.0000
6	1.0000	0.9992	0.9971	0.9946	0.9942	0.9965	0.9985	0.9999
7	1.0000	0.9993	0.9974	0.9954	0.9954	0.9975	0.9990	0.9999
8	1.0000	0.9993	0.9974	0.9954	0.9954	0.9975	0.9990	0.9999
9	1.0000	0.9994	0.9981	0.9971	0.9979	0.9991	0.9997	1.0000
10	1.0000	0.9994	0.9981	0.9971	0.9979	0.9991	0.9997	1.0000
11	0.9985	0.9975	0.9965	0.9956	0.9955	0.9970	0.9987	0.9999
12	0.9965	0.9924	0.9871	0.9808	0.9744	0.9746	0.9814	0.9977
13	0.9797	0.9603	0.9376	0.9135	0.8891	0.8766	0.8701	0.8648
14	0.9968	0.9991	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000
15	0.9976	0.9978	0.9981	0.9984	0.9991	0.9996	0.9998	1.0000
16	0.9994	0.9991	0.9991	0.9994	0.9997	0.9999	1.0000	1.0000
17	0.9989	0.9996	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000
18	0.9984	0.9982	0.9983	0.9986	0.9992	0.9997	0.9999	1.0000
19	0.8623	0.8813	0.9145	0.9509	0.9822	0.9939	0.9981	0.9999
20	0.9957	0.9949	0.9942	0.9937	0.9939	0.9959	0.9982	0.9999

† Genetic models from Table 3.9.

In the presence of both linkage and epistasis (lower half of Tables 3.12 and 3.13 except for the first column), the pattern is complicated. Model #13 (F_2 ratio of 15:1, and shows that as intensity of coupling linkage increases, the R^2 value increases up to unity, but as the intensity of repulsion linkage increases the R^2 value decreases. This demonstrates probable linkage-epistasis interaction.

To understand the interaction, results given in Table 3.5 for the general expression of $Cov(k, 1; n, n')$ were used to calculate a numerical example for Cov(2,1;2,3), the first covariance in Gates *et al.* (1960). The result is presented in Table 3.14 with linkage values $\lambda = 0.0$, 0.5 and 1.0 for model #13. Note that different signs may balance out linkage effects, so total epistatic variances may be the same for both $\lambda = 0.0$ and $\lambda = 1.0$.

From Tables 3.12 and 3.13, another general feature is that repulsion linkage causes more reduction in R^2 values than coupling linkage. In other words, for the same genetic model, repulsion linkage rejects the null hypothesis more frequently than coupling linkage. This is confirmed from estimation of probability of rejecting the null hypothesis, as shown Tables 3.15 and 3.16 for Model #2 (9:6:1 F_2 ratio) and Model #13 (15:1 F_2 ratio), respectively. The results from Tables 3.15 and 3.16 also reproduce the pattern shown in Tables 3.12 and 3.13. For Model #13 with the absence of linkage ($\lambda = 0$), probabilities of rejecting the null hypothesis are much higher (>50%) than the expected ($\leq$ 5%) since $F_{0.05}$ is used as the criterion. This confirms that epistasis acts like linkage.

It is clear from Tables 3.15 and 3.16 that the ratio, c, of the standard error of an estimated covariance to its expected value should be less than

Table 3.14: Epistatic components calculated for $Cov(2,1;2,3)$ for three different linkage values λ and Model #13 (15:1 F_2 ratio)

Epistatic component†	$\lambda = 1 - 2p$		
	0.0000	0.5000	1.0000
i^2_{ab}	0.2756	0.2501	0.1334
$j^2_{ab} + j^2_{ba}$	0.2668	0.2002	0.0000
l^2_{ab}	0.0500	0.0716	0.1334
$d_a j_{ab} + d_b j_{ba}$	-0.4270	-0.3002	0.0000
$\pm(h_a + h_b)\, i_{ab}$	0.0000	0.1668	0.2668
$(h_a + h_b)\, l_{ab}$	-0.1001	-0.1355	-0.2668
$\pm i_{ab} l_{ab}$	0.0000	-0.0636	-0.2668

† see Table 3.5.

0.0625 to give sufficiently high probability of detecting coupling linkagea and less than 0.125 to assure detection of repulsion linkages. This observation provides a basis for designing experiments to estimate variances and covariances (see Section 3.3).

The question of why repulsion linkage is easier to detect than coupling linkage can be considered as follows. Suppose that in a completely inbred population, there are only four homozygotes with respect to two loci A and B, $AABB$, $AAbb$, $aaBB$ and $aabb$. The frequencies of these genotypes for

Table 3.15: Estimated probabilities of rejecting the null hypothesis that there is no linkage using Gates' test calculated for Model #2 (9:6:1 F_2 ratio) with different linkages

	$\lambda = 1 - 2p$						
		Coupling			Repulsion		
c†	0.0	0.2	0.6	0.9	0.2	0.6	0.9
0.25000	0.038	0.047	0.052	0.036	0.068	0.327	0.453
0.12500	0.054	0.081	0.119	0.049	0.132	0.906	0.984
0.06250	0.043	0.301	0.538	0.073	0.575	1.000	1.000
0.03125	0.047	0.932	0.997	0.281	0.999	1.000	1.000

† Precision factor (see Section 3.2.1.3)

Table 3.16: Estimated probabilities of rejecting the null hypothesis that there is no linkage using Gates' test calculated for Model #13 (15:1 F_2 ratio) with different linkages

	$\lambda = 1 - 2p$						
		Coupling			Repulsion		
c†	0.0	0.2	0.6	0.9	0.2	0.6	0.9
0.25000	0.506	0.179	0.060	0.042	0.836	0.993	1.000
0.12500	0.983	0.638	0.066	0.046	1.000	1.000	1.000
0.06250	1.000	0.999	0.091	0.039	1.000	1.000	1.000
0.03125	1.000	1.000	0.356	0.046	1.000	1.000	1.000

† Precision factor (see Section 3.2.1.3)

both coupling and repulsion linkage and genotypic values are shown in Eqn.(3.25)

Genotype	Frequency at ∞		Genotypic value
	Coupling	Repulsion	
$AABB$	$\frac{1}{2(1+2p)}$	$\frac{p}{(1+2p)}$	$2x$
$AAbb$	$\frac{p}{(1+2p)}$	$\frac{1}{2(1+2p)}$	x
$aaBB$	$\frac{p}{(1+2p)}$	$\frac{1}{2(1+2p)}$	x
$aabb$	$\frac{1}{2(1+2p)}$	$\frac{p}{(1+2p)}$	0

(3.25)

Note that the genetic model is strictly additive. Then, $Cov(\infty,1;\infty,\infty)$ can be calculated for both coupling and repulsion linkages. For coupling linkage,

$$Cov(p;\infty,1;\infty,\infty) = x^2(\frac{1}{1+2p})$$

and for repulsion linkage,

$$Cov(-p;\infty,1;\infty,\infty) = x^2(\frac{2p}{1+2p}).$$

It is apparent that

$$\frac{Cov(p;\infty,1;\infty,\infty)}{Cov(-p;\infty,1;\infty,\infty)} = \frac{1}{2p}. \tag{3.26}$$

Unless $p = 0.5$, the ratio is always greater than unity. This proves a

well-known fact that coupling linkage gives rise to greater variances and covariances than repulsion linkage. It is then expected that given a set of 30 covariances, coupling linkage causes greater variability among these covariances, thus, higher R^2 values. The evidence from one numerical example supports such expectation. With a linkage value $\lambda = 0.6$, a few statistics of a set of 30 covariances $Cov(k,k';n,n')$ for Model #1 are shown in Table 3.17.

Table 3.17: Mean, standar deviation, minimum and maximum values of 30 covariances for Model #1 (1:4:6:4:1 F_2 ratio) with linkage $\lambda = 0.6$

λ†	Mean	St.dev.	Min.	Max.
+0.6	1.769	0.727	0.067	2.725
0.0	1.156	0.492	0.063	1.875
−0.6	0.544	0.268	0.058	1.025

† coupling (+) or repulsion (−) linkage.

The range for coupling linkage is almost twice as large as that for repulsion linkage. This simple example may be extended to any model.

In general, very high R^2 values were obtained even for the case of moderate coupling and repulsion linkages. Very little difference in the R^2 values between the presence and the absence of linkages suggests a difficulty in detecting linkage using Gates' test. Why are the high R^2 values obtained in the presence of linkage?

Recall that Gates *et al.* (1960) used a set of 30 generation numbers for k, k', n and n' in $Cov(k,k';n,n')$. In terms of linkage, k and k' are of special importance since they reflect the number of rounds of recombination occurring at gametogenesis in different generations of selfing. In order to facilitate discussion, a particular case where $k' = k-1$ is investigated.

The general expression for $Cov(k,k-1;n,n')$ with assumption of no epistasis can be obtained by substituting $k-1$ for k' in Eqn.(3.23). Thus,

$$Cov(k,k-1;n,n') = (\frac{1}{2})^{k-1} D_{k-1} + (\frac{1}{2})^{n+n'-k} H_{k-2}, \tag{3.27}$$

where $k-1$ = rank (Mather and Jinks, 1982, p192), and

$$D_{k-1} = (d_a^2 + d_b^2) \pm 2\lambda^{k-1} d_a d_b$$
$$H_{k-2} = (h_a^2 + h_b^2) + 2\lambda^2 \zeta^{k-2} h_a h_b$$

It is seen from Eqn.(3.27) that cross products of additive and dominance effects enter the covariances in the presence of linkage ($\lambda > 0$) and that these cross products decrease geometrically as the rank ($k-1$) increases arithmetically. Table 3.18 shows that except for D values of lower ranks (D_1 and D_2) or tight linkage ($\lambda \geq 0.8$) there are no essential differences between D values over the ranks and D values converge rapidly to that of no linkage. Note that calculations shown in Table 3.18 take the form $d_a = d_b = 1$. In Gates' test for linkage, values of k in 30 covariances range from 2 to 5, suggesting four different ranks involved. The average variability among these 30 covariances due to even moderate linkage is expected to be a small component of the total genetic variance. Thus, R^2 values usually are high even with the presence of linkage.

Table 3.18: Numerical values of additive variance D_r for $r = 1$ to 5 and $r = \infty$ with different values λ of coupling and repulsion linkages

	$\lambda = 1 - 2p$						
D_r†	0.800	0.500	0.200	0.000	0.200	0.500	0.800
	-------Coupling------				------Repulsion------		
D_1	3.600	3.000	2.400	2.000	1.600	1.000	0.400
D_2	3.280	2.500	2.080	2.000	1.920	1.500	0.720
D_3	3.024	2.250	2.016	2.000	1.984	1.750	0.976
D_4	2.819	2.125	2.003	2.000	1.997	1.875	1.181
D_5	2.656	2.063	2.001	2.000	1.999	1.938	1.344
D_∞	2.000	2.000	2.000	2.000	2.000	2.000	2.000

† D_r is an additive variance for rank r.

3.3. Statistical Considerations

3.3.1. Estimation and Sample Size

The analysis of variance and covariance model used in Gates' test is hierarchical. The total covariance, $Cov(k,k';n,n')$, can be subdivided into the components of families. When $k'=1$, each plant in F_k represents a family. When $k'=k-1$, a group of plants in F_n derived from the same F_{k-1} plant represents a family. The analysis of variance and covariance for $k'=k-1$ is more informative than that for $k'=1$. If a complete pedigree (From F_2 to F_k) of all progenies is known, then a complete specification of the components of covariance $Cov(k,k';n,n')$ can be achieved from the hierarchical analysis of covariances.

The question to be considered is as follows: For a given genetic situation (linkage and genic effect model) how large should an experiment be to ensure that the standard error of the estimated $\hat{Cov}(k,k';n,n')$, from the analysis of covariance, is not in excess of a constant, c, times its expectation, i.e., $s.e.(\hat{Cov}) < c \times \hat{Cov}$, or $s.e.(\hat{Cov})/\hat{Cov} < c$? The procedure and the argument involved will be given only for the two cases of lower order hierarchy, i.e. $k'=1$ and $k'=k-1$. The extension to any higher order of hierarchy should be apparent although more complicated.

3.3.1.1. Two-way classification ($k' = 1$)

Suppose that two random samples of progenies are taken from the F_n and $F_{n'}$ generation, both derived from a F_k plant by selfing. Data on those progenies from a replicated field trial with a randomized completed block design allows the following analysis of covariances (see Table 3.19) as well as analyses of variance for the two generations. It is assumed that the measured trait for these two progeny generations have a bivariate normal distribution. Thus, the marginal distribution for each generation is a univariate normal distribution.

Table 3.19: Form of analyses of variance or covariance for two-way classification ($k' = 1$)

Source	df	MS&MP†	EMS&MP†
Replicates	$r-1$	-	-
Progenies	$a-1$	m_1	$\sigma_e^2 + rCov(k,1;n,n')$
Error	$(a-1)(r-1)$	m_2	σ_e^2

† Mean squares (MS) and Mean products (MP).

The values of a, number of families in F_n and $F_{n'}$, and r, the number of replicates can be chosen so that the size of experiment $T = ar$ is minimum, the minimum size of the experiment required to ensure that $s.e.(\hat{Cov})/\hat{Cov} < c$. From Table 3.19,

$$\hat{Cov} = \frac{(m_1 - m_2)}{r}$$

and

$$s.e.(\hat{Cov}) = \sqrt{\frac{1}{r^2}[\sigma^2_{m_1} + \sigma^2_{m_2}]} \tag{3.28}$$

For large a values,

$$s.e.(\hat{Cov}) \simeq \sqrt{\frac{1}{r^2}[\frac{2m_1^2}{a-1} + \frac{2m_2^2}{(a-1)(r-1)}]},$$

where m_1 and m_2 can be found according to the least-squares principles.

$$m_1 = \hat{\sigma}^2_e + r\hat{Cov}$$
$$m_2 = \hat{\sigma}^2_e,$$

but $\hat{Cov}$ can be expressed as

$$\hat{Cov} = (\frac{h^2}{1-h^2})\hat{\sigma}^2_e = y\hat{\sigma}^2_e,$$

where $y = (\frac{h^2}{1-h^2})$ and h^2 is the heritability,

$$h^2 = \frac{Cov}{Cov + \sigma^2_e}.$$

Thus,

$$\frac{s.e.(\hat{Cov})}{\hat{Cov}} \simeq \frac{1}{yr}\sqrt{\frac{2(ry+1)^2}{a-1} + \frac{2}{(a-1)(r-1)}} < c. \tag{3.29}$$

Simplification of this expression yields

$$a - 1 > \frac{2[(ry+1)^2 - y(ry+2)]}{rc^2y^2(r-1)},$$

which can be substituted into $T = ar$,

$$T = r + \frac{2[(ry+1)^2 - y(ry+2)]}{c^2 y^2 (r-1)}.$$

For T to have a minimum value, the conditions where $dT/dr = 0$ and $d^2T/dr^2 > 0$ must hold. dT/dr and d^2T/dr^2 are the first and second order derivatives, respectively.

After a few steps of algebra, $dT/dr = 0$ yields

$$r > 1 \pm \frac{1}{y}\sqrt{\frac{2}{2+c^2}}$$

and $\frac{d^2T}{dr^2} > 0$ for $r > 1$. In order for T to have the minimum value,

$$r > 1 + \frac{1}{y}\sqrt{\frac{2}{2+c^2}} = 1 + \frac{d}{y}, \tag{3.30}$$

where $d = \sqrt{\frac{2}{2+c^2}}$ and thus,

$$a > \frac{2[(d+1)^2 + dy]}{c^2 d(y+d)} + 1 \tag{3.31}$$

Note that r and a are the functions of the heritability, h^2 and constant c, but the relations are in a complicated fashion.

3.3.1.2. Nested classification($k' = k-1$)

Now consider two random samples of progenies in the F_n and $F_{n'}$ generations derived from a group of F_k plants by selfing, but tracing to a single F_{k-1} plant. The analysis of covariances becomes more extensive (Table 3.20). Here a, b and r are numbers of F_k plants (families), F_k-derived F_n or F_k-dereived $F_{n'}$ lines (sub-families) and replicates, respectively.

Table 3.20: Form of analyses of variance or covariance for nested classification ($k' = k-1$)

source	df	MS&MP†	EMS&MP†
Replicates	$r-1$	-	-
Among F_{k-1}	$a-1$	m_1	$\sigma_e^2 + rCov' + brCov''$
Among F_k			
Within F_{k-1}	$a(b-1)$	m_2	$\sigma_e^2 + rCov'$
Error	$(ab-1)(r-1)$	m_3	σ_e^2

† Mean squares (MS) and Mean products (MP).

From Table 3.20,

$$\hat{Cov}'' = \frac{1}{br}(m_1 - m_2)$$

$$\hat{Cov}' = \frac{1}{r}(m_2 - m_3).$$

In terms of Gates' notation, $Cov' = Cov(k, k-1; n, n')$ and $Cov'' = Cov(k-1, 1; n, n')$. Thus, $Cov = Cov' + Cov'' = Cov(k, 1; n, n')$. Since Cov and Cov'' are often expressed as a linear function of genetic variance components, the less rigorous definitions for broad sense heritability may be expressed as: $h^2 = \frac{Cov}{Cov + \sigma_e^2}$ and $h''^2 = \frac{Cov''}{Cov'' + \sigma_e^2}$.

Let $x = \frac{h^2}{1-h^2}$ and $x'' = \frac{h''^2}{1-h''^2}$. then,

$$s.e.(\hat{Cov}'') = \frac{1}{br}\sqrt{\sigma^2_{m_1} + \sigma^2_{m_2}}$$

$$\simeq \frac{1}{br}\sqrt{\frac{2m_1^2}{a-1} + \frac{2m_2^2}{a(b-1)}} \tag{3.32}$$

$$= \frac{1}{br}\sqrt{\frac{2[1+r(x-x'')+brx'']^2}{a-1} + \frac{2[1+r(x-x'')]^2}{a(b-1)}}$$

thus,

$$\frac{s.e.(\hat{Cov}'')}{\hat{Cov}''} \simeq \frac{\frac{1}{br}\sqrt{\frac{2[1+r(x-x'')+brx'']^2}{a-1} + \frac{2[1+r(x-x'')]^2}{a(b-1)}}}{x''} < c \tag{3.33}$$

Similarly,

$$\frac{s.e.(\hat{Cov}')}{Cov'} \simeq \frac{\frac{1}{r}\sqrt{\frac{2[1+r(x-x'')]^2}{a(b-1)} + \frac{2}{(ab-1)(r-1)}}}{(x-x'')} < c. \tag{3.34}$$

The values of a, b and r can be therefore chosen such that the size of the experiment $T = abr$ is minimum.

In this case, it is difficult to obtain the analytic solution for a, b and r which satisfy the requirement. Thus, numerical evaluation is necessary. A PASCAL program was written for this purpose. Input parameters included constant c = 0.0625, 0.125; h^2 = 0.2, 0.4, 0.6 and 0.8 and h''^2 = 0.1, 0.3, 0.5 and 0.7. The values for h^2 and h''^2 cover the range from low to high heritabilities.

3.3.2. Results and discussion

3.3.2.1. Two-way classification ($k' = 1$)

In this simplest case, number of families a, and number of replicates r are the only two factors in determining the size of experiment. The two factors in turn depend on the heritability h^2 and the 'constant' c. Here h^2 is the measure of importance of environmental effects relative to genetic effects and c specifies the desired precision of the estimates of covariances. A set of values for r and a for different combinations of values of h^2 and c is presented in Table 3.21.

Table 3.21: Numbers of families(a) and replicates(r) which give rise to the mimimum size of the experiment for estimation of covariances between the selfed relatives

	c = 0.0625†		c = 0.125	
h^2	r	a	r	a
0.10	10	1897	10	475
0.30	3	1774	3	444
0.50	2	1281	2	321
0.70	2	973	2	244
0.90	2	669	2	168

† precision factor (see Section 3.2.1.3)

With both $c = 0.125$ and $c = 0.0625$, the number of replicates stays

around two even in the moderate h^2 (= 0.5), but a substantial increase in number of families (e.g., 168 to 321 for c = 0.125) takes place as h^2 decreases (from 0.9 to 0.5) for the same r. Note that when c reduces by one-half (e.g. from 0.125 to 0.0625), a must be quadrupled but r remains the same. Similarly, number of families a increases by 77 as h^2 decreases by 0.2 for c = 0.125 and 154 for c = 0.0625.

The large number of replicates (3 to 10) for low heritability (0.30 to 0.10) was obtained from statistical calculation. The total number of progenies required for evaluation in such an experiment precludes the possiblity of performing this experiment. Thus, in conducting the experiment, the number of replicates may be reduced to as few as two. This will lead to considerable reduction in the size of the experiment.

The number of families required for estimation of the covariances is at least 300 with moderate heritability (h^2 = 0.5) for the standard error of the estimates being about 12% as large as the estimates themselves (c = 0.125) (Table 3.21). This requirement must meet in order to detect the repulsion linkage (Tables 3.15 and 3.16). Detection of coupling linkage requires even higher precision (c = 0.0625) for the same magnitudes of heritability. This will increase the number of families to four times that required for c = 0.125. Thus, the number of families (94) used by Gates *et al.* (1960) is not sufficient for estimating $Cov(k,1;n,n')$, even if h^2 in the traits of their experiment is unity. Later, Croissant and Torrie (1970), VandeLogt (1978) and VandeLogt *et al.* (1984) used only 49 families to estimate $Cov(k,1;n,n')$. This number is much lower than required from the above calculation. In these later studies, estimates of $Cov(k,1;n,n')$ must have very

large sampling variability, and extremely low precision may have been obtained for the estimates of $Cov(k,1;n,n')$. It appears that such estimates are of limited value in determining the importance of linkage.

3.3.2.2. Nested classification ($k' = k - 1$)

Table 3.22 gives values of number of families, a, number of sub-families, b and number of replicates which give rise to the minimum size of experiment for eight different combinations of c, h^2 and h''^2. It is indicated from Table 3.22 that the number of families is again the most important factor in determining the size of experiment. However, the pattern produced by this table is less straightforward than that by Table 3.21. From low to moderate heritabilities (h^2 = 0.2 to 0.6) the number of families required is decreasing, the number of sub-families is reduced from 6 to 4 while the number of replicates is reduced from 3 to 2. With high heritability (h^2 = 0.8), the numbers of sub-families and replicates are reduced to the mimimum value of two but the number of families is increased, compared to that of the moderate heritability.

The total number of progenies ($a \times b$) required for estimation of the covariances with a given precision decreases as the heritability increases (Table 3.22). However, even with high heritability, the number of progenies is on the order of several hundreds. To allocate these progenies into a single complete block or replicate, one needs more land area. Increase in land area generally increases environmental variance. This in turn reduces the reliability of estimates of the covariances. A solution to this paradox is to reduce the size of block, using incomplete blocks or reduction in progeny numbers (treatments) (Cockerham, 1963).

Table 3.22: Numbers of families(a), sub-families(b) and replicates(r) which give rise to the mimimum size of the experiment for estimation of the covariances for different combinations of heritabilities h^2 and h''^2 and constant c

		c = 0.0625†			c = 0.125		
h^2	h''^2	a	b	r	a	b	r
0.2	0.1	1546	6	3	387	6	3
0.4	0.3	1108	4	3	277	4	3
0.6	0.5	811	4	2	203	4	2
0.8	0.7	1208	2	2	303	2	2

† precision factor (see Section 3.2.1.3)

The number of families plays the most important role in determining the precision of the estimated covariances. Statistical calculation suggest that more than two sub-families and replicates may be needed for low heritability (Table 3.22). With high heritability, two replicates are sufficient for detecting both coupling and repulsion linkages.

For most quantitative traits, low or moderate heritabilities may have to be used in planning the experiments if there is little *a priori* knowledge about heritability of the trait. Then two sub-families per family and two replicates may be used. These numbers were used by Gates *et al.* (1960), Croissant and Torrie (1971), VandeLogt (1978) and VandeLogt *et al.* (1984). However, these authors used many fewer families (49 families only in the three later studies) than the minimum required for detecting linkage.

4. Experimental Estimation of Genetic and Environmental Variances

4.1. Materials and methods

4.1.1. Genetic materials

Three crosses of spring wheat, Potam × Ingal, Ken 149 × HY320 and RL4137 × Ingal, were used in this study. Because of greater variability expressed for kernel hardness, heading date, plant height, spikelets/spike, kernels/spike, kernels/spikelet and grain yield, these three crosses were chosen among ten crosses made by Dr. R.J. Baker .

4.1.2. Mating design

4.1.2.1. Experiment I

A hierarchical scheme (Fig. 4.1) was used by Dr. Baker in developing materials of the three crosses. In each cross, there were 48 F_2-derived families. From each of these F_2-derived families, three F_3 plants were randomly selected to produce three F_3-derived F_4 lines. Seed from plants of each F_3-derived F_4 line was bulked to produce an F_3-derived F_5 line. In 1986, these F_3-derived F_5 lines were evaluated and the bulked seeds from each of these lines were used to produce F_3-derived F_6 lines. In 1987, the F_3-derived F_6 lines were evaluated.

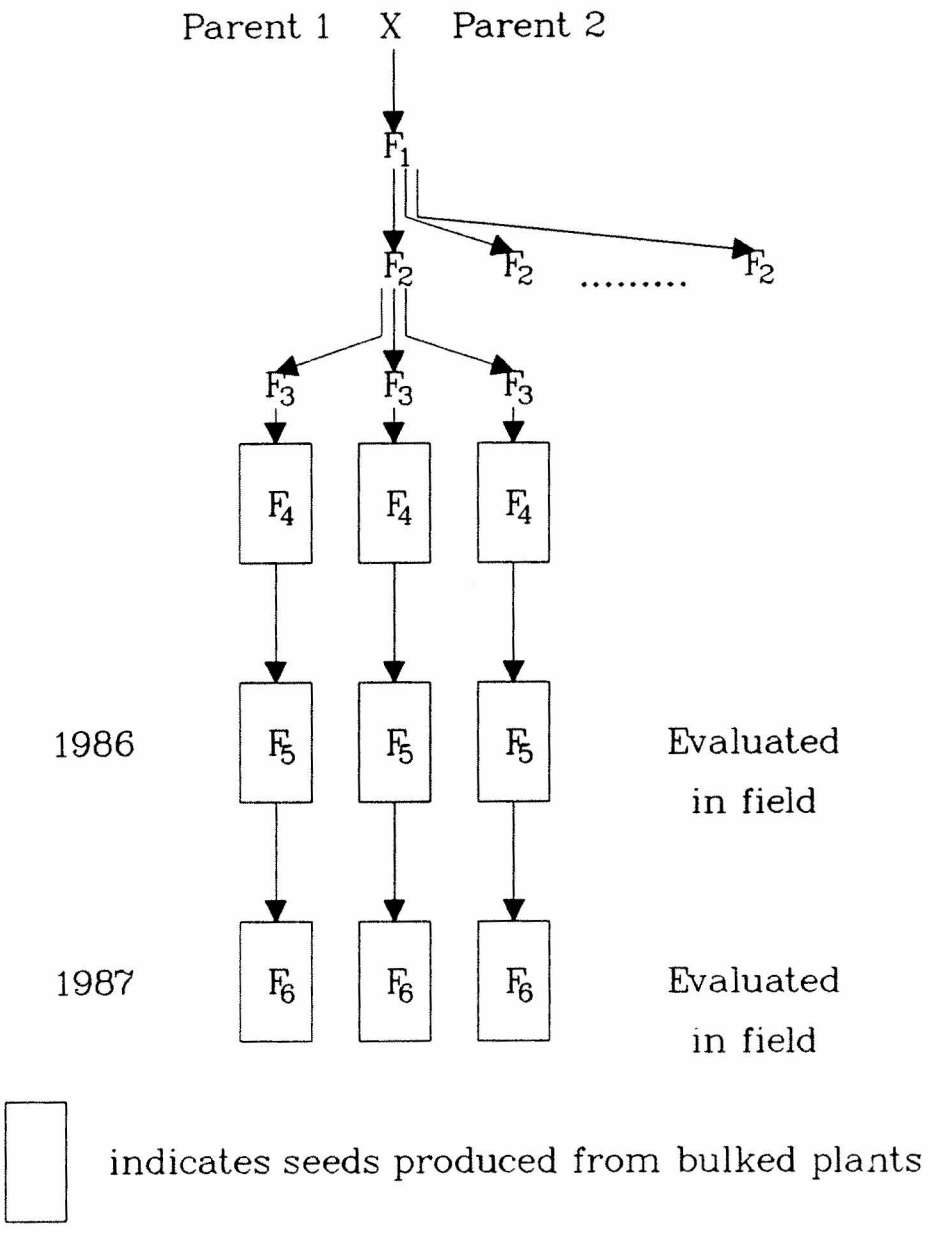

Figure 4.1: The hierarchical structure developed for Experiment I.

4.1.2.2. Experiment II

A more extensive hierarchical scheme (Fig. 4.2) was also used to develop another set of genetic materials of the three crosses. In each cross, there were 48 F_2-derived families. From each of these families, two F_3 plants were randomly selected to produce two F_2-derived F_3 lines. From each of these F_2-derived F_3 lines, two F_4 plants were randomly selected to produce two F_3-derived F_4 lines. From each of these F_3-derived F_4 lines lines, two plants were randomly selected to produce two F_4-derived F_5 lines. In 1986, seed from each of two randomly selected plants from each of these F_4-derived F_5 lines was used to produce F_4-derived F_6 lines which were evaluated in 1987.

4.1.3. Experimental design

4.1.3.1. Experiment I

A field experiment for genetic materials of Experiment I (Section 4.1.2.1) was conducted in a randomized complete block design with two replicates. For each replicate, 48 F_2-derived families were divided into four groups of 12 families with $12 \times 3 = 36$ F_3-derived lines per group. Randomization was performed for groups in each replicate and for the F_3-derived lines in each group. This experimental design was used in both 1986 and 1987.

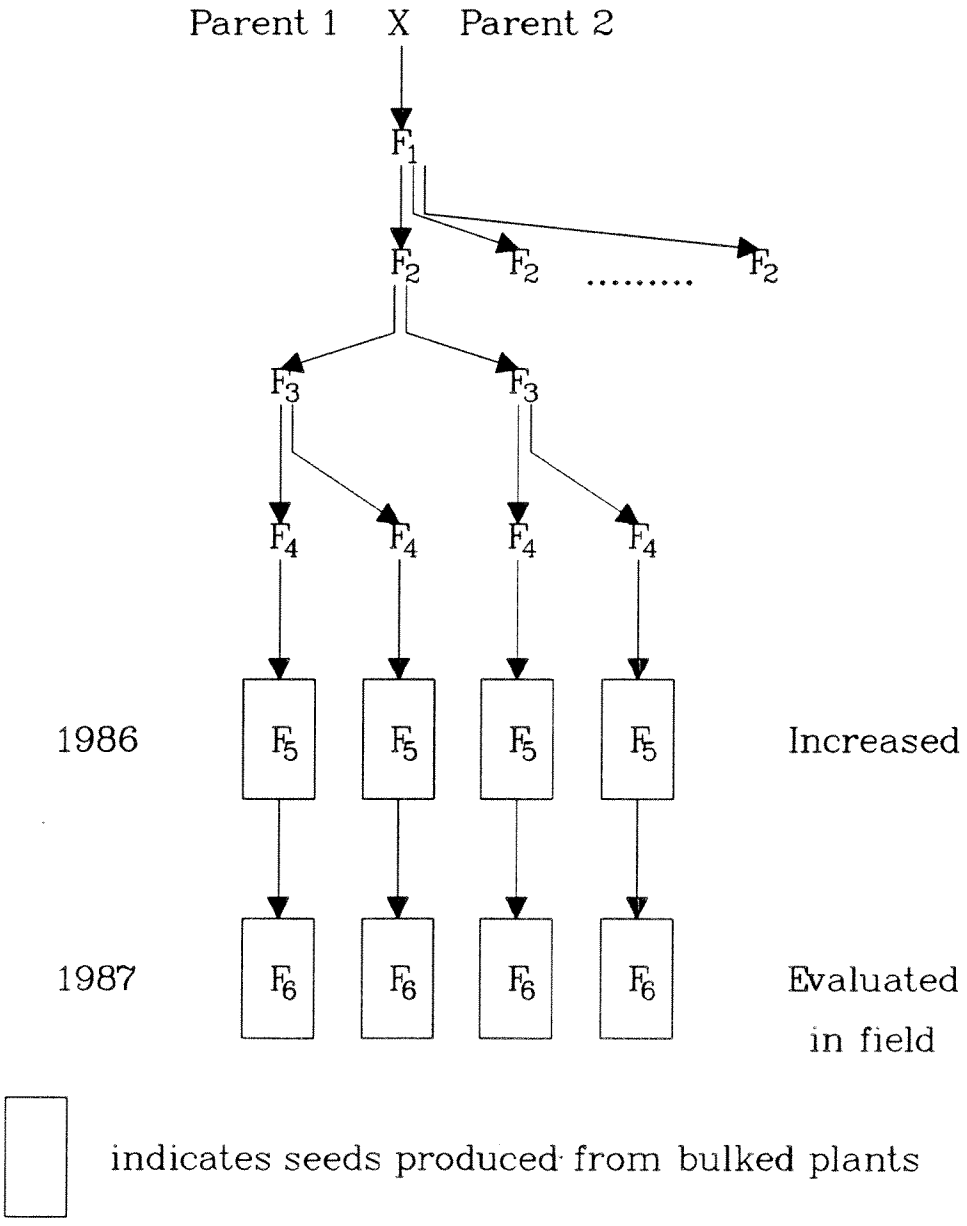

Figure 4.2: The hierarchical structure developed for Experiment II.

4.1.3.2. Experiment II

An unreplicated augmented design was used in 1987 to conduct a field experiment for the genetic materials of Experiment II (Section 4.1.2.2). Forty-eight F_2-derived families were divided into four groups of 12 families with $12 \times 2 \times 2 = 48$ F_4-derived F_6 lines per group. One arbitrarily chosen genotype (Katepwa) was used as the check. This check genotype was added 12 times in each group and was randomized along with the F_4-derived F_6 lines in that group. The replicated check genotype in each group provide an estimate of error variance for this unreplicated design.

4.1.4. Experimental procedures

4.1.4.1. Planting and harvesting

In both 1986 and 1987, the experiments were located at the University of Saskatchewan seed farm in Saskatoon. Each plot was a 2 m long single row with 30 cm spacing between rows. One hundred seeds were sown in each plot with seedplaced fertilizer (11-51-0, 56 kg/ha). For each experiment, four border plots at each side were added to alleviate border effects. Prior to harvest, a random sample of ten spikes from each plot was made for data collection.

4.1.4.2. Data collection

Data collection was carried out for Experiment I in both 1986 and 1987 and for Experiment II in 1987 only. Because of delayed maturity in some lines, materials from the cross Ken 149 × HY320 were not included in the study in 1987. Table 4.1 contains a list of abbreviations and descriptions of eight traits evaluated for each experiment. Kernel hardness was tested by a

Brabender micro hardness-tester calibrated with a Neepawa standard at 31 seconds.

4.1.5. Analysis of variance and covariance

4.1.5.1. Experiment I

An analysis of variance and covariance for all traits in each cross of Experiment I (Section 4.1.2.1) was computed using the following model:

$$X_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_{jk} + \delta_{jkl} + \epsilon_{ijkl}, \qquad (4.1)$$

where X_{ijkl} is the observation in the l*th* line in the the k*th* family in the j*th* group of the i*th* replicate, μ is population mean, α_i is the effect of the i*th* replicate, β_j is the effect of the j*th* group, $(\alpha\beta)_{ij}$ is the replicate × group interaction, γ_{jk} is the effect of the k*th* family nested in the j*th* group, δ_{jkl} is the effect of the l*th* line nested in the k*th* family and the j*th* group and ϵ_{ijkl} is the experimental error.

The expected mean squares for the analysis are given in Table 4.2. Note that replicates, groups in replicates, families in groups and lines in families are all considered random factors. Using terminology of Gates *et al.* (1960), components of variance for 1986 can be interpreted as σ^2_γ = *Cov*(2,1;5,5) and σ^2_δ = *Cov*(3,2;5,5). In 1987, σ^2_γ = *Cov*(2,1;6,6) and σ^2_δ = *Cov*(3,2;6,6).

It is possible to combine the data from both 1986 and 1987 to obtain covariances between the two years for Experiment I. Totals, over two replicates, of each F_3-derived F_5 line in 1986 were taken as one variable while totals for F_3-derived F_6 lines in 1987 were taken as the other. Then,

Table 4.1: Abbreviations and descriptions of measurements for traits evaluated in progeny of three spring wheat crosses

Abbreviation	Trait	Description
HD	Heading date	Calendar days from planting to date when 50% of the spikes were emerged from the flag leaf sheath
HT	Plant height	The distance in centimeters(cm) from ground to tip , excluding awns, as determined by an average of two measurements per plot
SPLSP	Spikelets/spike	Mean spikelets/spike of ten spikes selected at random from each plot
KLSP	Kernels/spike	Mean kernels/spike of ten spikes selected at random from each plot
KLSPL	Kernels/spikelet	Ratio of kernels/spike to spikelets/ spike
SPW	Spike yield	Mean weight in grams(g) of kernels of ten spikes selected at random from each plot
KW	Kernel weight	Total kernel weight in micrograms(mg) divided by total kernels of ten spikes
LKH	Kernel hardness	Time in seconds(sec) of grinding six grams of wheat kernels expressed on a logarithmic scale

Table 4.2: Form of analysis of variance and expected mean squares for a hierarchical model where replicates, groups, families and lines are all random (Experiment I)

Source	df	Expected mean squares
Replicates(R)	1	$\sigma^2_\epsilon + 36\sigma^2_{\alpha\beta} + 144\sigma^2_\alpha$
Groups(G)	3	$\sigma^2_\epsilon + 2\sigma^2_\delta + 6\sigma^2_\gamma + 36\sigma^2_{\alpha\beta} + 72\sigma^2_\beta$
R × G	3	$\sigma^2_\epsilon + 36\sigma^2_{\alpha\beta}$
F_2 Families	44	$\sigma^2_\epsilon + 2\sigma^2_\delta + 6\sigma^2_\gamma$
F_3 within F_2	96	$\sigma^2_\epsilon + 2\sigma^2_\delta$
Error	140	σ^2_ϵ

an analysis of covariance was used to partition the total sum of products into components due to covariation among groups, among F_2-derived families in groups, and among F_3-derived lines within families.

4.1.5.2. Experiment II

Statistical analysis of Experiment II was in two parts. First, an one-way analysis for the check genotype was used to estimate environmental variance (see Analysis A of Table 4.3). Secondly, the hierarchical structure of the genetic materials can be analyzed in the similar way to that of Experiment I (see Analysis B of Table 4.3). The analysis of variance model for Analysis A is

$$X_{ij} = \mu + \beta_i + \epsilon_{ij}, \tag{4.2}$$

and the analysis of variance model for Analysis B is

$$X_{ijkl} = \mu + \beta_i + \gamma_{ij} + \delta_{ijk} + \nu_{ijkl} + \epsilon_{ijkl}, \qquad (4.3)$$

where X, μ, β, γ, δ and ϵ have the same meanings as those in Eqn(4.1) and ν_{ijkl} is the effect of the l*th* F_4 line in the k*th* F_3 line in the j*th* F_2 family in the i*th* group.

Table 4.3: Form of analysis of variance and expected mean squares for Experiment II

Source	df	Expected mean squares
		Analysis A for check genotype
Among groups	3	$\sigma_\epsilon^2 + 12\sigma_\beta^2$
Error	44	σ_ϵ^2
		Analysis B for hierarchical structure
Among groups	3	$\sigma_\epsilon^2 + \sigma_\nu^2 + 2\sigma_\delta^2 + 4\sigma_\gamma^2 + 48\sigma_\beta^2$
F_2 families	44	$\sigma_\epsilon^2 + \sigma_\nu^2 + 2\sigma_\delta^2 + 4\sigma_\gamma^2$
F_3 within F_2	48	$\sigma_\epsilon^2 + \sigma_\nu^2 + 2\sigma_\delta^2$
F_4 within F_3	96	$\sigma_\epsilon^2 + \sigma_\nu^2$

In Experiment II, σ_γ^2, σ_δ^2 and σ_ν^2 are estimates of $Cov(2,1;6,6)$, $Cov(3,2;6,6)$ and $Cov(4,3;6,6)$, respectively.

4.1.6. Estimation of genetic and environmental variances

From the analysis of variance and covariance of each experiment as outlined in the previous section, mean squares and products from various sources (e.g., families, sub-families, etc.) were calculated. These observed mean squares and products can be expressed as the sum of expected mean squares and products and a random error. In matrix form,

$$\boldsymbol{m} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{e}, \tag{4.4}$$

where $\boldsymbol{m}$ is an (n × 1) vector of the observed mean squares, $\boldsymbol{X}$ is an (n × k) matrix of known coefficients determined by genetic structure and experimental design (with n ≥ k), $\boldsymbol{b}$ is a (k × 1) vector of genetic and environmental variances to be estimated, and $\boldsymbol{e}$ is an (n × 1) vector of error random variables with mean vector $E(\boldsymbol{e}) = \boldsymbol{0}$ and variance-covariance matrix $Var(\boldsymbol{e}) = E(\boldsymbol{e}\boldsymbol{e}') = \Sigma$.

In order to consider both correlated and unequal errors of the observed mean squares and products, two methods of estimation, weighted least squares (WLS) of Nelder (1960) and maximum likelihood (ML) of Hayman (1960), were used in the present study to estimate the genetic and environmental variances under the three alternative genetic models specified below:

$$\begin{aligned}
&\text{Model I} && \boldsymbol{b} = (\sigma^2_{\epsilon}, \sigma^2_{A}, \sigma^2_{AA})' \\
&\text{Model II} && \boldsymbol{b} = (\sigma^2_{\epsilon}, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5, \beta_6)' \\
&\text{Model III} && \boldsymbol{b} = (\sigma^2_{\epsilon 1}, \sigma^2_{\epsilon 2}, \sigma^2_{A1}, \sigma^2_{A2}, \sigma_{A1A2})'
\end{aligned} \tag{4.5}$$

In Model I, the genetic variance components to be estimated were

limited to additive, σ^2_A, and additive × additive epistatic, σ^2_{AA}, variances. The common error variance, σ^2_ϵ, was included in this model to suggest homogeneity of the observed error variances between the two years. Note also that no dominance and no dominance types of epistasis were involved in this model, and the model assumes no linkage.

Model II is the model from Gates' test for linkage except that no attempt was made to estimate β_5 and β_6 because of zero coefficients for these two quadratic components. Once again the common error variance was included in this model.

Model III was used to allow for estimation of heterogeneous error and additive genetic variances between the two years, where $\sigma^2_{\epsilon 1}$ and $\sigma^2_{\epsilon 2}$ are the error variance for 1986 and 1987, σ^2_{A1} and σ^2_{A2} are the additive variances for both years and σ_{A1A2} is the covariance of the additive affects between the two years. In a general sense, this model is considered to be the genotype-environmental interaction model.

The expectations of the observed mean squares and products of interest can be obtained from a combined analysis of variance and covariance set forth in Table 4.4 for Experiment I and in Table 4.5 (a short form of Table 4.3) for Experiment II. For example, the F_2 family mean square of Experiment I in 1986 is

$$E(m_1) = \sigma^2_\epsilon + 2Cov(3,2;5,5) + 6Cov(2,1;5,5)\,.$$

For Model I, the expected genetic composition of $Cov(3,2;5,5)$ is $0.25\sigma^2_A$ + $0.3125\sigma^2_{AA}$, and the expectation of $Cov(2,1;5,5)$ is $0.5\sigma^2_A$ + $0.25\sigma^2_{AA}$. Thus, under Model I,

Table 4.4: Combined analysis of variance and covariance for Experiment I

Source	df	Mean squares	Expected mean squares†
			1986
F_2 families	44	m_1	c3 + 2c2 + 6c1
F_3 within F_2	96	m_2	c3 + 2c2
Error	140	m_3	c3
			1987
F_2 families	44	m_4	c6 + 2c5 + 6c4
F_3 within F_2	96	m_5	c6 + 2c5
Error	140	m_6	c6
			1986-87
F_2 families	44	m_7	2c8 + 6c7
F_3 within F_2	96	m_8	2c8

† c1 = Cov(2,1;5,5), c2 = Cov(3,2;5,5), c4 = Cov(2,1;6,6), c5 = Cov(3,2;6,6), c7 = Cov(2,1;5,6), c8 = Cov(3,2;5,6), and c3, c6 are error mean squares $\sigma^2_{\epsilon 1}$ and $\sigma^2_{\epsilon 2}$for each year for Model III and become equal for Models I and II.

$$E(m_1) = \sigma^2_\epsilon + 3.5\sigma^2_A + 2.125\sigma^2_{AA}.$$

The coefficients of expectations, i.e. the $\boldsymbol{X}$ matrices, with the three models are presented in Table 4.6 for Experiment I and in Table 4.7 for Experiment II. Note that only Model I was obtained for Experiment II since there were only four observed mean squares available for estimation. Both WLS and

Table 4.5: A short form of Table 4.3 with pertinent part for estimation

Source	df	Mean squares	Expected mean squares†
			1987
F_2 families	44	m_9	σ_ϵ^2 + c11 + 2c10 + 4c9
F_3 within F_2	48	m_{10}	σ_ϵ^2 + c11 + 2c10
F_4 within F_3	96	m_{11}	σ_ϵ^2 + c11
Error	44	m_{12}	σ_ϵ^2

†c9 = Cov(2,1;6,6), c10 = Cov(3,2;6,6), c11 = Cov(4,3;6,6).

Table 4.6: Matrices of coefficients relating the components of genetic variance to the statistics observed in Experiment I under three models

Observed	Model I			Model II					Model III				
statistic	σ^2_ϵ	σ^2_A	σ^2_{AA}	σ^2_ϵ	β_1	β_2	β_3	β_4	$\sigma^2_{\epsilon 1}$	$\sigma^2_{\epsilon 2}$	σ^2_{A1}	σ^2_{A2}	σ_{A1A2}
m_1	1	3.5	2.125	1	7	0.1562	-2	-0.0626	1	0	3.5	0.0	0.0
m_2	1	0.5	0.625	1	1	0.0626	-2	-0.0626	1	0	0.5	0.0	0.0
m_3	1	0.0	0.000	1	0	0.0000	0	0.0000	1	0	0.0	0.0	0.0
m_4	1	3.5	2.125	1	7	0.0390	-2	-0.0156	0	1	0.0	3.5	0.0
m_5	1	0.5	0.625	1	1	0.0156	-2	-0.0156	0	1	0.0	0.5	0.0
m_6	1	0.0	0.000	1	0	0.0000	0	0.0000	0	1	0.0	0.0	0.0
m_7	0	3.5	2.125	0	7	0.0780	-2	-0.0312	0	0	0.0	0.0	3.5
m_8	0	0.5	0.625	0	1	0.0312	-2	-0.0312	0	0	0.0	0.0	0.5

Table 4.7: Matrix of coefficients relating the components of genetic variances to the statistics observed in Experiment II

Observed statistic	σ^2_ϵ	σ^2_A	σ^2_{AA}
m_9	1	2.625	1.82813
m_{10}	1	0.625	0.82813
m_{11}	1	0.125	0.20313
m_{12}	1	0.000	0.00000

ML methods require calculation of the weight matrix, which is the inverse of the error variance-covariance matrix, Σ. The Σ matrix can be obtained by calculating various sampling variances and covariances of the observed mean squares and products (for WLS) and those of the expected mean squares (for ML). The general formulas for these sampling variances and covariances can be found in Kendall and Stuart (1963, p235). Specifically, suppose that in a bivariate normally distributed population, M_{20}, M_{02} and M_{11} are the observed (or expected) mean squares and mean products and represent the estimates of σ^2_1, σ^2_2 and $\rho\sigma_1\sigma_2$, respectively. Then various sampling variances and covariances of interest are found below:

$$Var(M_{20}) = \frac{2M_{20}^2}{df} \qquad Var(M_{02}) = \frac{2M_{02}^2}{df}$$

$$Var(M_{11}) = \frac{M_{11}^2 + M_{20}M_{02}}{df}$$

$$Cov(M_{20}, M_{11}) = \frac{2M_{20}M_{11}}{df} \qquad Cov(M_{02}, M_{11}) = \frac{2M_{02}M_{11}}{df}$$

$$Cov(M_{20}, M_{02}) = \frac{2M_{11}^2}{df}$$

For Experiment I, a bivariate normal distribution for both mean squares and products among F_2 families and within F_2 families of the two years were assumed. Thus, an 8 × 8 Σ matrix is given by

$$
\begin{bmatrix}
\frac{2m_1^2}{df_1} & 0 & 0 & \frac{2m_7^2}{df_1} & 0 & 0 & \frac{2m_1m_7}{df_1} & 0 \\
0 & \frac{2m_2^2}{df_2} & 0 & 0 & \frac{2m_8^2}{df_2} & 0 & 0 & \frac{2m_2m_8}{df_2} \\
0 & 0 & \frac{2m_3^2}{df_3} & 0 & 0 & 0 & 0 & 0 \\
\frac{2m_7^2}{df_1} & 0 & 0 & \frac{2m_4^2}{df_4} & 0 & 0 & \frac{2m_4m_7}{df_4} & 0 \\
0 & \frac{2m_8^2}{df_2} & 0 & 0 & \frac{2m_5^2}{df_5} & 0 & 0 & \frac{2m_5m_8}{df_5} \\
0 & 0 & 0 & 0 & 0 & \frac{2m_6^2}{df_6} & 0 & 0 \\
\frac{2m_1m_7}{df_1} & 0 & 0 & \frac{2m_4m_7}{df_4} & 0 & 0 & \frac{m_7^2+m_1m_4}{df_7} & 0 \\
0 & \frac{2m_2m_8}{df_2} & 0 & 0 & \frac{2m_5m_8}{df_5} & 0 & 0 & \frac{m_8^2+m_2m_5}{df_8}
\end{bmatrix}
$$

Here m_i are given in Table 4.4 and $df_1 = df_4 = df_7 = 44$, $df_2 = df_5 = df_8 = 96$ and $df_3 = df_6 = 140$.

With the ML method, the Σ is calculated by use of the expected mean squares $\boldsymbol{X}\hat{\boldsymbol{b}}$, where

$$\hat{\boldsymbol{b}} = (\boldsymbol{X}'\Sigma^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\Sigma^{-1}\boldsymbol{m} \tag{4.6}$$

is only given implicitly. The explicit solution is obtained iteratively by calculating Σ initially from the observed mean squares and products $\boldsymbol{m}$, finding a first estimate $\hat{\boldsymbol{b}}_1$, recalculating Σ from $\boldsymbol{X}\hat{\boldsymbol{b}}_1$, and finding a second estimate $\hat{\boldsymbol{b}}_2$, and so on. The iterations are continued until the χ^2 reaches a stable minimum (Hayman, 1960). The χ^2 is given by

$$\boldsymbol{m}'\Sigma^{-1}\boldsymbol{m} - \boldsymbol{b}'\boldsymbol{X}'\Sigma^{-1}\boldsymbol{m} \tag{4.7}$$

The χ^2 values for testing goodness-of-fit were recorded for each case of the two methods of estimation.

For Experiment II, a 4 × 4 Σ matrix can be obtained in the same way as that for Experiment I except that no covariances are involved in this Σ matrix. Thus, the weights of this diagonal matrix may be obtained simply as the reciprocals of the sampling variances.

4.2. Results and discussion

The climatological data are presented in Appendix C for monthly mean temperature (°C), monthly total precipitation (mm) and monthly total evaporation (mm) from May to August. 1987 summer was warmer and much drier than the average (1951-80) and 1986 summer.

4.2.1. Analysis of variance and covariance

The mean squares from the analysis of Experiment I for each of eight traits are presented in Table 4.8 for the cross Potam × Ingal, in Table 4.9 for cross the RL4137 × Ingal and in Table 4.10 for the cross Ken 149 × HY320 (1986 only). With one exception, variation among F_2-derived families and among F_3-derived lines within F_2 families was significant at $P = 0.01$ for all eight traits.

With few exceptions, blocking (replicate) and grouping had little effect on reducing the error variance for each trait. This was indicated by the lack of significant variation among replicates and groups. However, there was a significant replicate × group interaction for most traits of each cross. This suggests that grouping was effective in reducing environmental variation. The coefficients of variation for all traits were less than 10 percent.

The pertinent part of the results from the analysis of Experiment II for both crosses Potam × Ingal and RL4137 × Ingal are shown in Table 4.11. A similar pattern to that of Experiment I was observed. However, in a few cases, variances among F_4-derived lines within F_3 were less than the corresponding error variances. This may reflect the fact that genetic variation within a highly inbred ancestral population becomes quite small.

Table 4.8: Mean squares from analyses of variance for eight traits measured on progeny lines derived from an F_2 population of Potam x Ingal (Experiment I)

Source	df	Trait†							
		HD	HT	SPLSP	KLSP	KLSPL	SPW	KW	LKH
					1986				
Replicates(R)	1	34.03	6810.42*	10.35	673.75	1.722	0.0226	439.56*	5.160
Groups(G)	3	22.48	421.43	3.07	67.84	0.157	0.2113	62.22	0.357
R x G	3	7.51**	412.22**	1.25**	76.61**	0.239**	0.0285*	19.57**	1.957
F_2 families	44	10.44**	224.71**	4.70**	107.52**	0.288**	0.2006**	56.14**	1.137**
F_3 within F_2	96	1.75**	33.04**	0.87**	14.96**	0.041*	0.0184**	14.67**	0.236**
Error	140	0.41	8.56	0.19	6.65	0.030	0.0077	2.99	0.079
					1987				
Replicates(R)	1	7.03	44.34	7.28	148.49	0.170	0.0654	16.53	1.866
Groups(G)	3	10.20	74.25	3.84	64.28	0.283	0.0598	55.69	0.333
R x G	3	1.11	21.96	1.84	19.20	0.114*	0.0163	14.32**	0.055
F_2 families	44	5.11**	113.59**	5.19**	37.86**	0.150**	0.0680**	29.38**	0.938**
F_3 within F_2	96	1.61**	27.25**	0.75**	14.90**	0.050**	0.0149**	9.01**	0.133**
Error	140	0.58	14.19	0.29	7.30	0.030	0.0087	2.60	0.029

† See Table 4.1 for full descriptions.

*,** Significant at P = 0.05 and 0.01, respectively.

Table 4.9: Mean squares from analyses of variance for eight traits measured on progeny lines derived from an F_2 population of RL4137 x Ingal (Experiment I)

Source	df	Trait†							
		HD	HT	SPLSP	KLSP	KLSPL	SPW	KW	LKH
					1986				
Replicates(R)	1	0.17	25.98	0.92	13.26	0.170	0.0512	15.26	0.0015
Groups(G)	3	9.11	130.83	1.72	117.86	0.631	0.1204	12.88	0.0400
R x G	3	1.48*	63.93**	1.33**	29.38**	0.115**	0.0141*	3.51*	0.0043
F_2 families	44	10.72**	61.72**	1.59**	45.18**	0.142**	0.0421**	21.47**	0.0395**
F_3 within F_2	96	2.84**	22.09**	0.58**	9.59**	0.039**	0.0117**	7.24**	0.0082**
Error	140	0.49	10.65	0.19	3.77	0.018	0.0038	1.14	0.0016
					1987				
Replicates(R)	1	2.72	1192.35	0.02	295.65	1.345*	0.0397	108.41	0.0085
Groups(G)	3	22.17	66.26	5.42	163.70	0.604	0.1840	14.37	0.0325
R x G	3	5.38**	28.97*	0.10	32.75**	0.109*	0.0285*	11.73**	0.0103
F_2 families	44	6.15**	67.50**	2.66**	36.12**	0.097**	0.0295**	15.54**	0.0676**
F_3 within F_2	96	1.51**	27.97**	0.75**	15.42**	0.052**	0.0159**	5.48**	0.0149**
Error	140	0.46	7.53	0.26	7.68	0.032	0.0074	1.65	0.0029

† See Table 4.1 for full descriptions.

*,** Significant at P = 0.05 and 0.01, respectively.

Table 4.10: Mean squares from analyses of variance for eight traits measured on progeny lines derived from an F_2 population of Ken 149 x HY320 (Experiment I)

Source	df	Trait†							
		HD	HT	SPLSP	KLSP	KLSPL	SPW	KW	LKH
					1986				
Replicates(R)	1	12.50	663.09	0.45	31.97	0.230	0.001	61.00	0.2447
Groups(G)	3	33.04	516.55	3.34	66.06	0.099	0.075	14.81	1.2031
R x G	3	9.69**	1181.22**	2.55**	100.80**	0.230**	0.130**	42.13**	0.0215
F_2 families	44	29.20**	161.87**	5.19**	51.57**	0.117**	0.134**	41.82**	1.0130**
F_3 within F_2	96	4.99**	28.66**	1.38**	19.97**	0.044**	0.047**	13.77**	0.0765**
Error	139(140)‡	0.60	8.57	0.33	4.64	0.017	0.013	3.26	0.0089

† See Table 4.1 for full descriptions.

‡ The degrees of freedom in parenthesis are for heading date and height, and there is one missing value for the remaining traits.

*,** Significant at P = 0.05 and 0.01, respectively.

Table 4.11: Mean squares from analyses of variance for eight traits measured on progeny lines derived from an F_2 population of two crosses (Experiment II)

Source	df	Trait†							
		HD	HT	SPLSP	KLSP	KLSPL	SPW	KW	LKH
				Potam x Ingal					
F_2 families	44	3.46**	79.61*	3.45**	38.79*	0.148**	5.44**	35.32**	0.8310**
F_3 within F_2	48	1.31*	40.42**	0.80**	20.56**	0.066*	2.08*	12.72**	0.1186*
F_4 within F_3	96	0.81	16.78	0.30**	9.53*	0.040*	1.19*	5.25**	0.0724**
Error	44	0.87	23.06	0.12	5.62	0.025	0.73	1.46	0.0061
				RL4137 x Ingal					
F_2 families	44	6.31**	64.84*	2.04*	23.55*	0.080*	2.72*	15.62**	0.0822**
F_3 within F_2	48	2.31**	34.02**	1.02**	13.70*	0.049*	1.44*	7.00*	0.0171**
F_4 within F_3	96	0.90	12.20	0.48**	7.62	0.029	0.78	4.39**	0.0067*
Error	44	0.58	12.44	0.18	6.08	0.027	0.79	1.60	0.0039

† See Table 4.1 for full descriptions.

*,** Significant at P = 0.05 and 0.01, respectively.

Table 4.12: Mean products from analyses of covariance for eight traits measured on progeny lines derived from an F_2 population of two crosses (Experiment I)

Source	df	Trait† HD	HT	SPLSP	KLSP	KLSPL	SPW	KW	LKH
				Potam x Ingal					
Groups(G)	3	12.712	138.743	1.870	-23.393	-0.203	0.0883	41.995	-0.278
F_2 families	44	6.162**	141.547**	4.376**	49.272**	0.110**	0.1003**	32.583**	0.976**
F_3 within F_2	96	1.198**	17.497**	0.486**	7.170**	0.015**	0.0074**	7.513**	0.129**
				RL4137 x Ingal					
Groups(G)	3	11.053	-19.721	-0.183	33.056	0.085	0.0770	2.557	0.0333
F_2 families	44	7.089**	56.708**	1.582**	21.798**	0.038*	0.0212**	16.008**	0.0458**
F_3 within F_2	96	1.582**	18.184**	0.371**	5.518**	0.019**	0.0066**	4.445**	0.0077**

† See Table 4.1 for full descriptions.

*,** Significant at P = 0.05 and 0.01, respectively.

The results from analyses of covariance between two years for eight traits are presented in Table 4.12 for Experiment I. According to tests of correlation coefficients, calculated as ratios of covariances between two years to the geometric mean of variances for both years, covariation among F_2 families and among F_3-derived lines within F_2 families was significant at P = 0.01 for all the traits.

In Experiment I, variances and covariances among F_2 families and among F_3 lines within F_2 families were generally larger in cross the Potam × Ingal than in the cross RL4137 × Ingal, except for heading date.

4.2.2. Estimation of genetic and environmental variances

With limited number of estimates of mean squares and products, the best approach is to fit them to several alternative models and then to test for goodness-of-fit of the observed mean squares and products to the models. In the present study, three alternative models (as described in the previous section) were used. The χ^2 values for testing goodness-of-fit are presented in Table 4.13 for the WLS method and in Table 4.14 for the ML method. With both WLS and ML methods, the χ^2 values for additive and additive × additive model (Model I) and Gates' model (Model II) were significantly high, suggesting that there are significant deviations of the observed mean squares and products from these two models. With few exceptions (particularly in cross RL4137 × Ingal), the χ^2 values for Model III showed that there was no significant departure of the observed mean squares and products from the model. The fit of the observed mean squares and products to Model III suggests that either error or additive genetic variances or both are heterogeneous between the two years.

Table 4.13: Chi-square for goodness-of-fit of the observed mean squares and products to their expectations by weighted least squares method for three models with eight traits in two crosses

Model	df	Trait†							
		HD	HT	SPLSP	KLSP	KLSPL	SPW	KW	LKH
					Potam x Ingal				
Model I	5	17.1**	22.7**	18.1**	14.4**	13.0**	17.4**	23.0**	43.0**
Model II	3	5.1	8.8*	13.7**	2.7	7.5	1.5	4.4	25.1**
Model III	3	6.5	0.7	3.5	6.4	7.5	7.2	3.3	1.7
					RL4137 x Ingal				
Model I	5	28.2**	10.2	16.4**	35.3**	25.0**	24.9**	16.6**	32.3**
Model II	3	2.1	4.3	16.0**	31.2**	21.9**	21.0**	5.8	27.1**
Model III	3	4.2	12.5**	3.9	3.6	6.5	6.2	7.3	0.9

† See Table 4.1 for full descriptions.

*,** indicates significant deviations of the observed mean squares from their expectations at P = 0.05 and 0.01, respectively.

Table 4.14: Chi-square for goodness-of-fit of the observed mean squares and products to their expectations by maximum likelihood method for three models with eight traits in two crosses

Model	df	Trait†							
		HD	HT	SPLSP	KLSP	KLSPL	SPW	KW	LKH
					Potam x Ingal				
Model I	5	106.6**	68.4**	30.5**	115.7**	100.6**	173.3**	105.5**	47.8**
Model II	3	5.5	8.9*	17.1**	4.1	27.9**	1.7	6.2	38.6**
Model III	3	5.7	0.7	3.9	6.8	7.0	10.0*	2.8	1.6
					RL4137 x Ingal				
Model I	5	85.1**	10.0	43.0**	78.1**	77.1**	47.9**	25.1**	100.4**
Model II	3	1.8	5.0	39.9**	42.9**	45.9**	22.9**	5.2	60.2**
Model III	3	3.6	10.2*	3.8	3.6	6.5	6.7	6.1	0.9

† See Table 4.1 for full descriptions.

*,** indicates significant deviations of the observed mean squares from their expectations at P = 0.05 and 0.01, respectively.

In deriving the expectations of the sample variances and covariances between relatives from selfing, Horner and Weber (1956) and Gates et al. (1960) did not consider genotype-environmental interaction. Since they grew different generations in different years genotype-environmental interaction may have been an important factor in determining sample variances and covariances. In fact, Gates *et al.* (1960), in detecting and estimating linkage for eight traits of a soybean cross, reanalyzed the same set of data by using the alternative model of genotype-environmental interaction and found that with the new analysis, only additive variance was significant for flowering time, flowering time to maturity and yield. This was in contrast to the results from analysis of data by using Gates' model with linkage where Gates *et al.* (1960) found linkage effects for flowering time and yield, in addition to significant additive and dominance effects. The new analysis by Gates *et al.* (1960) suggests the presence of genotype-environmental interaction for these three traits.

One might argue that the lack of fit of the observed mean squares and products to Gates' linkage model would be due to fewer sample variances and covariances obtained in the present study. However, in a selfing system, there is high correlation between generations (Brim and Cockerham, 1961). This is indicated by the fact that the coefficients of different genetic components are highly correlated for many sets of relatives in Gates' model. From a statistical point of view, high correlations between independent variables in regression analysis are called *collinearity* or *ill conditioning* (Belsley *et al.*, 1980). The presence of collinearity will reduce the reliability of the estimates because it induces large sampling variability for the estimates.

Belsley *et al.* (1980) provided a measure of collinearity given by

$$\kappa = \frac{\mu_{max}}{\mu_{min}},$$

where κ is called *condition number* and μ_{max} and μ_{min} are the largest and the smallest eigenvalues of $\boldsymbol{X}'\boldsymbol{X}$ with $\boldsymbol{X}$ being the matrix of coefficients derived from the expectations (see Table 4.6). For Models I, II and III, the κ values were found to be 521.029, 591357.720 and 9.064, respectively. Thus, Model III has the lowest κ value, while Gates' model has the highest κ value. A rule of thumb is that if the square root of κ is greater than, 30 then there is indication of severe collinearity (Weisberg, 1985). Thus, Gates' model (Model II) involves severe collinearity. Therefore, increasing number of sample variances and covariances between selfed relatives may be not useful in Gates' model for reducing the high χ^2 values. Model I is also not recommended because the condition number for this model is also much higher than that for Model III.

The high χ^2 values may also reflect the fact that the error variance-covariance matrix Σ of the observed mean squares and products is not always positive definite. One of the requirements of applying both WLS and ML methods is that the Σ matrix be positive definite; i.e., that all diagonal elements of this matrix be positive and that all eigenvalues be positive (Guttman, 1982). Although all diagonal elements may be positive, eigenvalues may not necessarily be so. The Σ matrix can be checked with regard to whether or not it is positive definite by calculating simple and partial correlations (Baker, 1986). If the partial correlation falls in the acceptable range of -1 to +1, then it is likely that the Σ matrix is positive

definite and can be used in calculating the weight matrix. In the present study, the 8 × 8 Σ matrix for Experiment I was positive definite.

Because Model III is likely to be the only fit model, estimation of error and genetic variances under this model was carried out by both WLS and ML methods. The WLS estimates from this model are presented in Table 4.15 for the cross Potam × Ingal and in Table 4.16 for the cross RL4137 × Ingal. The ML estimates are presented in Table 4.17 for the cross Potam × Ingal and in Table 4.18 for the cross RL4137 × Ingal. Both methods provided the similar estimates of genetic and environmental variances. Note that estimates of additive genetic variances for all eight traits in the two crosses were significantly different from their corresponding standard errors. This suggests that additive variance in these traits is of principal importance.

It has been expected that, in terms of the χ^2 test for goodness-of-fit, the ML method should be at least as good as the WLS method, since it involves additional computation. Recall from Tables 4.13 and 4.14 that both methods have χ^2 values which are similar in magnitude. In actual computation, it was noted that the χ^2 values stabilized after the second iteration of the ML method.

Two unweighted least squares procedures (Nasoetion *et al.*, 1967) were also compared with the WLS and ML methods. These two methods differ in whether the observed mean squares or the so-called design variance components (estimated from the observed mean squares) are used as dependent variables in the estimation. Neither method requires calculation of the weight matrix. The unweighted least squares estimates were similar to those estimates by the WLS and ML methods, but the standard errors of

Table 4.15: Weighted least squares estimates of error and additive genetic variances and their standard errors, genetic correlations and heritabilities of eight traits for Potam x Ingal (Experiment I)

Estimate or standard error	Trait†							
	HD	HT	SPLSP	KLSP	KLSPL	SPW	KW	LKH
$\hat{\sigma}^2_{\epsilon 1}$	0.3848	8.4162	0.1964	6.2345	0.0267	0.0076	3.0401	0.0795
s.e.$(\hat{\sigma}^2_{\epsilon 1})$	0.0476	0.9986	0.0222	0.7662	0.0034	0.0009	0.3491	0.0085
$\hat{\sigma}^2_{A1}$	2.6625	53.7035	1.1838	20.7827	0.0431	0.0263	18.5239	0.2839
s.e.$(\hat{\sigma}^2_{A1})$	0.3990	7.8597	0.1743	3.7836	0.0110	0.0050	2.6673	0.0429
$\hat{h}^2_1$	0.8737	0.8645	0.8577	0.7692	0.6175	0.7758	0.8590	0.7812
$\hat{\sigma}^2_{\epsilon 2}$	0.5663	14.1807	0.2858	7.6390	0.0300	0.0092	2.6853	0.0281
s.e.$(\hat{\sigma}^2_{\epsilon 2})$	0.0626	1.4966	0.0326	0.8005	0.0033	0.0010	0.2965	0.0033
$\hat{\sigma}^2_{A2}$	1.4975	25.9234	1.1018	8.6593	0.0350	0.0107	9.4344	0.2280
s.e.$(\hat{\sigma}^2_{A2})$	0.2405	4.6375	0.1788	1.8214	0.0077	0.0024	1.4627	0.0321
$\hat{h}^2_2$	0.7256	0.6464	0.7940	0.5313	0.5385	0.5377	0.7784	0.8903
$\hat{\sigma}_{A1A2}$	1.8786	35.9287	1.0466	11.9903	0.0263	0.0151	11.6534	0.2534
s.e.$(\hat{\sigma}_{A1A2})$	0.2805	5.3833	0.1562	2.2358	0.0068	0.0029	1.7428	0.0349
$\hat{r}_g$	0.9408	0.9629	0.9164	0.8938	0.6771	0.9001	0.8815	0.9960

† See Table 4.1 for full descriptions.

Table 4.16: Weighted least squares estimates of error and additive genetic variances and their standard errors, genetic correlations and heritabilities of eight traits for RL4137 x Ingal (Experiment I)

Estimate or standard error	Trait†							
	HD	HT	SPLSP	KLSP	KLSPL	SPW	KW	LKH
$\hat{\sigma}^2_{\epsilon 1}$	0.4871	9.0970	0.1985	3.6572	0.0175	0.0040	1.1803	0.0020
s.e.$(\hat{\sigma}^2_{\epsilon 1})$	0.0567	1.0510	0.0216	0.4365	0.0021	0.0005	0.1334	0.0002
$\hat{\sigma}^2_{A1}$	3.6283	18.6406	0.4937	11.6255	0.0357	0.0124	7.6219	0.0117
s.e.$(\hat{\sigma}^2_{A1})$	0.5055	3.2939	0.0823	2.0146	0.0071	0.0021	1.0995	0.0017
$\hat{h}^2_1$	0.8816	0.6720	0.7132	0.7607	0.6711	0.7561	0.8659	0.8540
$\hat{\sigma}^2_{\epsilon 2}$	0.4341	7.3455	0.2631	7.8534	0.0319	0.0073	1.7204	0.0030
s.e.$(\hat{\sigma}^2_{\epsilon 2})$	0.0500	0.8422	0.0296	0.8505	0.0034	0.0008	0.1837	0.0003
$\hat{\sigma}^2_{A2}$	1.8997	22.5889	0.8041	9.6883	0.0217	0.0079	5.0918	0.0210
s.e.$(\hat{\sigma}^2_{A2})$	0.2862	3.6192	0.1312	1.9925	0.0058	0.0017	0.8054	0.0029
$\hat{h}^2_2$	0.8140	0.7546	0.7535	0.5523	0.4049	0.5197	0.7475	0.8750
$\hat{\sigma}_{A1A2}$	2.4899	20.8624	0.5556	7.5387	0.0167	0.0077	5.9918	0.0146
s.e.$(\hat{\sigma}_{A1A2})$	0.3490	3.1801	0.0896	1.5752	0.0047	0.0015	0.8687	0.0020
$\hat{r}_g$	0.9484	1.0167	0.8818	0.7103	0.6000	0.7780	0.9618	0.9314

† See Table 4.1 for full descriptions.

Table 4.17: Maximum likelihood estimates of error and additive genetic variances and their standard errors, genetic correlations and heritabilities of eight traits for Potam x Ingal (Experiment I)

Estimate or standard error	Trait†							
	HD	HT	SPLSP	KLSP	KLSPL	SPW	KW	LKH
$\hat{\sigma}^2_{\epsilon 1}$	0.3914	8.4232	0.1983	6.3009	0.0275	0.0076	3.0389	0.0799
s.e.$(\hat{\sigma}^2_{\epsilon 1})$	0.0477	0.9988	0.0222	0.7673	0.0034	0.0009	0.3492	0.0085
$\hat{\sigma}^2_{A1}$	2.9338	54.8939	1.2527	24.0128	0.0564	0.0365	20.1855	0.2893
s.e.$(\hat{\sigma}^2_{A1})$	0.4138	7.9980	0.1781	4.0561	0.0121	0.0063	2.8597	0.0432
$\hat{h}^2_1$	0.8823	0.8670	0.8633	0.7921	0.6722	0.8277	0.8692	0.7836
$\hat{\sigma}^2_{\epsilon 2}$	0.5772	14.1901	0.2873	7.8432	0.0302	0.0092	2.6938	0.0282
s.e.$(\hat{\sigma}^2_{\epsilon 2})$	0.0627	1.4968	0.0326	0.8097	0.0033	0.0010	0.2969	0.0033
$\hat{\sigma}^2_{A2}$	1.6188	26.1890	1.1781	9.3295	0.0351	0.0125	10.4283	0.2359
s.e.$(\hat{\sigma}^2_{A2})$	0.2547	4.6601	0.1874	1.8551	0.0078	0.0026	1.5801	0.0327
$\hat{h}^2_2$	0.7372	0.6486	0.8039	0.5433	0.5375	0.5760	0.7947	0.8932
$\hat{\sigma}_{A1A2}$	2.0269	36.5088	1.1083	13.1868	0.0280	0.0194	12.9225	0.2599
s.e.$(\hat{\sigma}_{A1A2})$	0.2920	5.4432	0.1609	2.2880	0.0070	0.0034	1.8977	0.0353
$\hat{r}_g$	0.9301	0.9629	0.9123	0.8810	0.6293	0.9082	0.8907	0.9949

† See Table 4.1 for full descriptions.

Table 4.18: Maximum likelihood estimates of error and additive genetic variances and their standard errors, genetic correlations and heritabilities of eight traits for RL4137 x Ingal (Experiment I)

Estimate or standard error	Trait†							
	HD	HT	SPLSP	KLSP	KLSPL	SPW	KW	LKH
$\hat{\sigma}^2_{\epsilon 1}$	0.4882	9.6982	0.2007	3.6698	0.0176	0.0040	1.1860	0.0020
s.e.$(\hat{\sigma}^2_{\epsilon 1})$	0.0567	1.0892	0.0217	0.4367	0.0021	0.0005	0.1336	0.0002
$\hat{\sigma}^2_{A1}$	4.0024	23.9689	0.5689	12.3803	0.0411	0.0143	9.5749	0.0119
s.e.$(\hat{\sigma}^2_{A1})$	0.5507	3.8486	0.0914	2.0563	0.0074	0.0023	1.3365	0.0017
$\hat{h}^2_1$	0.8913	0.7119	0.7392	0.7714	0.7002	0.7814	0.8898	0.8561
$\hat{\sigma}^2_{\epsilon 2}$	0.4397	7.4939	0.2629	7.9228	0.0325	0.0075	1.7316	0.0030
s.e.$(\hat{\sigma}^2_{\epsilon 2})$	0.0503	0.8443	0.0296	0.8517	0.0035	0.0008	0.1845	0.0003
$\hat{\sigma}^2_{A2}$	2.0582	30.3071	0.8694	10.7952	0.0266	0.0105	6.1440	0.0216
s.e.$(\hat{\sigma}^2_{A2})$	0.2998	4.5935	0.1379	2.1131	0.0062	0.0020	0.9224	0.0030
$\hat{h}^2_2$	0.8240	0.8018	0.7678	0.5767	0.4501	0.5833	0.7801	0.8780
$\hat{\sigma}_{A1A2}$	2.7270	27.3088	0.6263	8.2442	0.0205	0.0100	7.4223	0.0150
s.e.$(\hat{\sigma}_{A1A2})$	0.3756	3.9638	0.0981	1.6504	0.0052	0.0018	1.0395	0.0021
$\hat{r}_g$	0.9501	1.0132	0.8905	0.7131	0.6200	0.8161	0.9677	0.9356

† See Table 4.1 for full descriptions.

the estimates by the unweighted methods were generally larger than those by the weighted methods. In addition, a negative estimate of error variance was found in the cross Potam × Ingal. Another disadvantage of the unweighted least squares procedures is that they cannot provide the χ^2 test for goodness-of-fit because the sum of squares of deviations of the observed variances from the expected values in the unweighted analysis is, by definition, not the χ^2.

The variance estimates for the same traits from separate years were quite different. Specifically, the estimates of additive variance were higher in 1986 than in 1987, while the estimates of environmental variance were lower in 1986 than in 1987 (particularly for the cross Potam × Ingal). This was indicated also by the estimates of heritabilities (Tables 4.15 to 4.18), measured by, for example,

$$\hat{h}_1^2 = \hat{\sigma}_{A1}^2/(\hat{\sigma}_{A1}^2 + \hat{\sigma}_{\epsilon 1}^2),$$

which were generally higher in 1986 than in 1987. However, genetic correlations (Tables 4.15 to 4.18), estimated by

$$\hat{\sigma}_{A1A2}/(\hat{\sigma}_{A1}\hat{\sigma}_{A2}),$$

were close to unity, suggesting that the additive genetic variation over the two years was similar. Confirmation of this result requires further analysis of heterogeneity.

To test for heterogeneity of the estimates over years, three reduced models (assuming Model III is a full model) were considered:

$$\text{Model IIIa} \quad \boldsymbol{b} = (\sigma^2_{\epsilon}, \sigma^2_{A1}, \sigma^2_{A2}, \sigma_{A1A2})'$$

$$\text{Model IIIb} \quad \boldsymbol{b} = (\sigma^2_{\epsilon 1}, \sigma^2_{\epsilon 2}, \sigma^2_{A})' \tag{4.8}$$

$$\text{Model IIIc} \quad \boldsymbol{b} = (\sigma^2_{\epsilon}, \sigma^2_{A})'$$

The χ^2 values were recorded for these three reduced models. The test of overall heterogeneity of estimates over the two years is provided by the difference in χ^2 values between Models III and IIIc. Similarly, the test of environmental heterogeneity is given by the difference between Models III and IIIa and the test of genetic heterogeneity by the difference between Models III and IIIb. Since both WLS and ML methods give similar results from these these tests, only those by the ML method are presented in Table 4.19 for the two crosses.

Note that since the genetic variances are not independent of the error variances in the estimation equations, the χ^2 tests of genetic and environmental heterogeneity do not necessarily sum to the overall heterogeneity χ^2. From Table 4.19 , the test of the overall heterogeneity was highly significant at P = 0.01 except for plant height of the cross RL4137 × Ingal. The test of environmental heterogeneity was also highly significant with two exceptions, while that of genetic heterogeneity was significant at P = 0.01 in only five out of 16 cases. Thus, the heterogeneity over years is likely due to heterogeneity of environmental variation, but not to change in genetic variances. An additional note from field experiment may provide supportive evidence to this conclusion. In 1987, seed germination was very poor while the germination in 1986 was normal (particularly for the cross

Table 4.19: Chi-square from analyses of genotype-environmental interaction (heterogeneity) over the two years in two crosses by the maximum likelihood method

Source of heterogeneity	df	Trait†							
		HD	HT	SPLSP	KLSP	KLSPL	SPW	KW	LKH
					Potam x Ingal				
Overall	3	97.2**	72.3**	27.6**	114.6**	101.7**	209.2**	99.0**	46.1**
Environmental	1	99.6**	74.1**	25.3**	110.1**	99.5**	191.9**	79.5**	4.6
Genetic	2	5.9	10.0**	4.4	1.9	0.2	1.3	0.5	36.7**
					RL4137 x Ingal				
Overall	3	79.5**	5.6	38.3**	74.3**	69.5**	43.0**	21.6**	98.1**
Environmental	1	66.8**	3.4	29.0**	65.4**	70.0**	37.4**	22.2**	68.8**
Genetic	2	0.4	3.3	2.4	21.1**	14.2**	13.0**	4.4	6.6*

† See Table 4.1 for full descriptions.

*,** indicates significant heterogeneity at P = 0.05 and 0.01, respectively

Potam × Ingal). This difference and the different climatological paterns (Appendix C) must increase environmental variation between the two years.

For Experiment II, because of the limited number of observed mean squares (only four), Model I was used. Although the χ^2 test for goodness of fit suggests that there is no significant departure of observed mean squares from the model, many negative estimates were found under this model by both WLS and ML methods. The ML method is supposed to always give positive estimates. This anomalous result for Experiment II indicates that Model I (additive and additive × additive) is not adequate. This is in agreement with the above analysis for Experiment I.

5. General Discussion and Conclusions

Gates' test for linkage is a procedure by which genetic variance and covariance among relatives from selfing are utilized to detect and estimate linkage of genes at different loci controlling a quantitative trait. The exact expectation of variances and covariances are derived with the assumption of no epistasis (Gates *et al.*, 1957). Moreover, Gates' test was developed from approximate expectations of variances and covariances (Gates et al., 1960). The approximate expectations become exact if the further assumption of no linkage is made.

Since both linkage and epistasis reflect interaction between genes at different loci, the detection and estimation of linkage from variances and covariances may be biased by the presence of epistasis. In the deterministic simulation carried out in the present study, four categories of genetic models were utilized to evaluate Gates' test for linkage:

(1) absence of both linkage and epistasis;

(2) presence of linkage and absence of epistasis;

(3) absence of linkage and presence of epistasis;

(4) presence of both linkage and epistasis.

For all four categories of models, the null hypothesis was that both linkage and epistasis are absent. Coefficients of determination, R^2, were calculated as the proportion of the total genetic variability among variances and

covariances that could be explained under the null hypothesis. Thus, models of the first category should give perfect fits and R^2 values of unity were expected. This was demonstrated by the results displayed in Tables 3.12 and 3.13. The R^2 values for the models of the remaining three categories should be less than unity and it is this expectation that provides the basis for Gates' test for linkage. However, this was not the case in many of these models (see Tables 3.12 and 3.13).

It may not be surprising to obtain R^2 values of unity or near unity for the models with very tight coupling and repulsion linkages ($\lambda > 0.9$) and the absence of epistasis since genes at different loci in these models act like those at a single locus. Some models of the fourth category (with the presence of both linkage and epistasis) have R^2 values of unity as well. This suggests that effects of linkage and epistasis may cancel each other out so that only additive and dominance effects are left. Linkage and epistasis can not be separated in variances and covariances and can interact in a very complicated fashion (see Table 3.5).

For models with the presence of epistasis but no linkage, R^2 values were also less than unity (Tables 3.12 and 3.13). The theoretical study (Section 3.1) indicates that epistasis may mimic the effect of linkage. Thus Gates' test for linkage fails to differentiate this model from that with the presence of linkage but with the absence of epistasis. In other words, epistasis, if present, may act like linkage in Gates' test for linkage.

For plant genetic materials with no *a priori* knowledge on gene action, the assumption of no epistasis is questionable. With complementary (9:7) interaction in the absence of linkage, Baker (1984) found that additive $\times$

additive epistatic variance in this case could be as large as 50 percent of additive variance. It would be desirable before performing Gates' test for linkage that there be a means of detecting whether or not epistasis is present.

Kearsey and Jinks (1968) proposed a triple test cross for detecting epistasis. By using the results in Section 3.1 on Hayman's device, validity of this triple test cross method can be readily checked. In the triple test cross, a random sample of F_2 individuals are used as pollen parents and crossed to three testers, the two inbred parents which differ in n ($n \geq 2$) segregating loci and the F_1 derived from the cross between these two parents. Progeny of crosses between an F_2 plant and each parent are denoted as L1 and L2, while that of the cross with the F_1 is denoted as L3. For $n = 2$, let $\Theta'(\theta_a', \theta_b')$ represent the genotype of each tester and let $\Theta''(\theta_a'', \theta_b'')$ be the genotype of an F_2 plant. If the presence of linkage has any impact on the triple test cross test for epistasis, then, with the assumption of no epistasis, expectation of (L1 + L2 - 2L3) will have a non-zero value. The expectations of progeny of crosses between the F_2 plants and the three testers with no epistasis derived from F_k-metric are given in Table 5.1

It is apparent from Table 5.1 that E(L1 + L2 - 2L3) = 0. This is also true for the two parents with genotypes *AAbb* and *aaBB*. Thus, the triple test cross is likely a valid test for epistasis, regardless of whether or not linkage is present. However, since epistatic deviations, like dominance deviations, may be either plus or minus and may cancel out when averaged over loci, the triple test cross may not be able to detect epistasis in some multi-locus cases.

Table 5.1: Expectations of progeny of crosses between the F_2 plants and the three testers in triple test cross

Tester	θ_a'	θ_b'	Expectation†
$AABB$	1	1	$E(L1)=\frac{1}{2}(1+\theta_a'')d_a+\frac{1}{2}(1+\theta_b'')d_b$ $+\frac{1}{2}(F_k-\theta_a'')h_a+\frac{1}{2}(F_k-\theta_b'')h_b$
$aabb$	−1	−1	$E(L2)=-\frac{1}{2}(1-\theta_a'')d_a-\frac{1}{2}(1-\theta_b'')d_b$ $+\frac{1}{2}(F_k+\theta_a'')h_a+\frac{1}{2}(F_k+\theta_b'')h_b$
$AaBb$	0	0	$E(L3)=\frac{1}{2}\theta_a''d_a+\frac{1}{2}\theta_b''d_b+\frac{1}{2}F_kh_a+\frac{1}{2}F_kh_b$

†Common factor m is omitted.

Genetic variances and covariances are functions of the squared genetic effects and, consequently, provide a more positive means of measuring the genetic effects, but estimates of variances and covariances have larger sampling variability than generation means. Hence, the triple test cross method is still recommended by Mather and Jinks (1982) for detecting epistasis.

In carrying out this test, crosses between F_2 plants and the three testers may not be able to produce sufficient seeds for progeny testing. For crops such as wheat which are natural self-fertilizers, but which produce little F_1 seed, Kearsey and Jinks (1968) pointed out that it is still possible to apply the method by using selfed progeny from the crosses.

Even if the assumption of no epistasis for Gates' test is satisfactory,

the R^2 values remain extremely high. Discussion on the high R^2 values was given in Section 3.2. The high R^2 values require that $Cov(k,k';n,n')$ must be estimated with high precision from an experiment. It was shown in Section 3.3 that the number of families is the single important factor in determining the precision of estimates of $Cov(k,k';n,n')$, while both number of replicates and/or number of lines per family can be minimized (about 2).

Recall that Gates' test for linkage has approximate expectations for variances and covariances while Mather's rank statistics method is an exact method. Bias due to the approximation was discussed in Section 3.1. The advantage of Gates' test over Mather's method is that it makes use of all types of variances and covariances, or $Cov(k,k';n,n')$ while Mather's method only uses the type of $Cov(k,k-1;n,n')$. $k-1$ in Mather's method is always referred to as the rank of the statistics. The difficulty of this method is that whenever the rank is increased by unity, two additional parameters are needed, one for additive variance and the other for dominance variance. Gates *et al.* (1960) modified Mather's method to develop a test which was approximate but had fewer parameters to be estimated. Both methods seem insensitive to variation among variances and covariances caused by linkage (see Section 3.2).

Cockerham's (1984b) method for detecting linkage is to use the probability of identity by descent of various combinations of alleles to determine the coefficients of quadratic components in the relevant covariances. For selfing, he makes the assumption that the initial population is in linkage equilibrium, and then uses changes in the additive × additive epistatic variance to detect the effect of linkage. This suggests that his

method is applicable only to the cross-fertilizing species where development of inbreds is a major task. The advantage of his method is the accomodation of multiple alleles. This, however, is at the expense of complication for any general formulation (Weir and Cockerham, 1977).

In the present study, Hayman's device was used to derive a general expression for covariance between relatives from selfing with epistasis and with the presence of both linkage and identity disequilibria. The advantage of Hayman's device over the traditional factorial model in the case of two alleles per locus is that it allows definition of homozygous effects rather than random mating additive effects. This immediately reduces the number of parameters in covariances. In addition, because gene and genotypic frequencies are functions of Hayman's θ-variable, Hayman's device is useful for linkage study.

In evaluating Gates' test for linkage by both theoretical analysis and computer simulation, the genetic model used in the present study was limited to two loci. Extension to more than two loci is straightforward. Thus, for n loci, the F_k-metric becomes

$$\begin{aligned} M(\theta_a, \theta_b, \theta_c, \ \ldots \ ; F_k) = m + \sum_a \theta_a d_a + \sum_a \left(\frac{1+F_k}{2} - \theta_a^2\right) h_a \\ + \sum_{a<b} \theta_a \theta_b i_{ab} + \sum_{a<b} \theta_a \left(\frac{1+F_k}{2} - \theta_b^2\right) j_{ab} \\ + \sum_{a<b} \left(\frac{1+F_k}{2} - \theta_a^2\right)\left(\frac{1+F_k}{2} - \theta_b^2\right) l_{ab} \\ + \sum_{a<b<c} \theta_a \theta_b \theta_c i_{abc} + \ \ldots \ . \end{aligned} \tag{5.1}$$

However, even with the absence of linkage, derivation of variance and covariance between relatives from selfing becomes much more complicated than in a two-locus case. Bulmer (1980, p69) gave the derivation of $Cov(k,k-1;n,n)$ with digenic interactions only in the absence of linkage. The complicated results from this derivation makes it has any practical value od the precedure doubtful. In self-fertilizing crops, it is reasonable to assume only additive types of epistatic variability because gene effects involving heterozygosis in these species are much less important than in cross-fertilizing species (Brim and Cockerham, 1961; Matzinger and Cockerham, 1963). This assumption simplifies the results considerably. Thus, the multi-locus $Cov(k,k-1;n,n')$ becomes

$$Cov(k,k-1;n,n') = (\frac{1}{2})^{k-1}\sum_a d_a^2 + (\frac{1}{2})^{n+n'-k}\sum_a h_a^2 + (\frac{1}{2})^{2k-2}(2^k-3)\sum_{a<b} i_{ab}^2 + \cdots . \quad (5.2)$$

Note that genic effects are defined with reference to the F_∞-metric.

With the presence of linkage for more than two loci, an additional parameter is required to specify a new linkage relation for every even number of loci (Jones, 1960; Schnell, 1961). The total number of linkage relations for n loci is $2^{n-1} - 1$.

In the two-locus case, the linkage contribution to additive genetic variance in variances and covariances between relatives from selfing could be either positive or negative, depending on whether the F_1 is in the coupling

or repulsion phase. This should also be true for the multi-locus case as long as the F_1 population is not a mixture of coupling and repulsion phases.

Among all possible F_1 genotypes resulting from four heterozygous loci, the proportion of pairwise repulsion linkages is about 55% and becomes 78% if six heterozygous loci are considered. This latter result can be found in Bliss and Gates (1968). However, the magnitude of the contribution due to coupling linkages, although less frequent, is greater than that of repulsion linkages (Bliss and Gates, 1966; Bulmer, 1980). Thus, for many linked loci, linkage contributions to additive variance tend to cancel out. Dominance variance involving linkage is independent of linkage phases but affected by the signs of dominance effects.

With linkage equilibrium, linkage effects are expressed only when there is epistasis (Cockerham, 1956). In self-fertilizing crops, the additive types of epistatic variances are important components. Thus coefficients associated with i^2_{ab} in Table 3.5 and those with higher order additive epistatic variances should be investigated.

In estimating genetic and environmental variances from field experiments by use of three alternative models, both weighted least squares and maximum likelihood estimates showed the lack of fit of the observed mean squares and products to Gates' linkage model. the genotype-environmental interaction model (Model III) was satisfactory. This may reflect the fact that Gates' model does not take into account genotype-environmental interaction which is possibly an important component in variances and covariances between relatives, in addition to linkage and epistasis. This may be particularly true when the estimates of variances and

covariances are obtained from relatives in different generations grown in different years. The data used by Gates *et al.* (1960) were collected from the generations grown in four different years, yet in the analysis of variance for separate years, the common error variance was used in the expected mean squares by Gates *et al.* (1960).

Later studies by Croissant and Torrie (1971), VandeLogt (1978) and VandeLogt *et al.* (1984) modified the experimental design so that different generations of the same families were grown in the same main plots and progeny lines were in subplots. The problem with this design is that it results in an environmental correlation between different generations within the same families. Sample variances and covariances between generations will be inflated by the environmental correlation (Falconer, 1981). In order to eliminate such correlations, one could choose to randomize lines and families within groups (see Section 4.1.3).

In the past, genotype-environmental interaction has been ignored by the studies of covariances between relatives. The main weaknesses of these studies are that if genotype-environmental interaction is important, then the estimates of variances and covariances become biased. The severity of biasedness depends on the nature of the interaction. Baker (1988a,b) argued that the genotype-environmental interaction is important only if it involves a change in rank (crossover interaction). However, scaling effects that do not result in changes in rank can still cause serious bias in estimation and interpretation of variances and covariances.

In conclusion, Gates' test for linkage fails to differentiate effects due to linkage from those due to epistasis. Even in the absence of epistasis, linkage

can not be easily detected by this test unless there is a strong repulsion linkage or a very large experiment. Gates' test has limited utility for detecting linkage of genes controlling a quantitative trait. Furthermore, in studying covariances between relatives, linkage, epistasis and genotype-environmental interaction must be all considered in order to obtain unbiased estimates of genetic variances. This is important since response to selection in any breeding program is a function of genetic variances.

6. References

Allard, R.W. 1960. Principles of plant breeding. Wiley, London.

Altman, D.W., and R.H. Busch. 1984. Random intermating before selection in spring wheat. Crop Sci. 24:1085-1089.

Bailey,Jr., T.B., and R.E. Comstock. 1976. Linkage and the synthesis of better genotypes in self-fertilizing species. Crop Sci. 16:363-370.

Baker, R.J. 1968. Extent of intermating in self-pollinated species necessary to counteract the effects of genetic drift. Crop Sci. 8:547-550.

Baker, R.J. 1978. Issues in diallel analysis. Crop Sci. 18:533-536.

Baker, R.J. 1984. Quantitative genetic principles in plant breeding. Pages 147-176 *in* J.P. Gustafson, ed. Gene manipulation in plant improvement. Plenum Press, New York.

Baker, R.J. 1986. Selection indices in plant breeding. CRC Press, Inc., Boca Raton, FL.

Baker, R.J. 1988a. Differential response to environmental stress. Pages 492-504 *in* B.S. Weir, E.J. Eisen, M.M. Goodman and G. Namkoong, eds. Proc. 2nd Intl. Conf. on Quantitative Genetics. Sinauer Assoc., Inc., Sunderland, MA.

Baker, R.J. 1988b. Tests for crossover genotype-environmental interactions. Can. J. Plant Sci. 68:405-410.

Belsley, D.A., E. Kuh, and R.E. Welsch. 1980. Regression diagnostics: Identifying influential data and sources of collinearity. Wiley, New York.

Bennett, J.H. 1954. The distribution of heterogeneity upon inbreeding. J. Roy. Statist. Soc. B16:88-99.

Bliss, F.A., and C.E. Gates. 1968. Directional selection in simulated

populations of self-pollinated plants. Aust. J. Biol. Sci. 21:705-719.

Bos, I. 1977. More arguments against intermating F2 plants of a self-fertilizing crop. Euphytica 26:33-46.

Bridges,Jr., W.C., and S.J. Knapp. 1987. Probabilities of negative estimates of genetic variances. Theor. Appl. Genet. 74:269-274.

Brim, C.A., and C.C. Cockerham. 1961. Inheritance of quantitative characters in soybeans. Crop Sci. 1:47-52.

Bulmer, M.G. 1980. The mathematical theory of quantitative genetics. Oxford University Press, Oxford, U.K.

Cockerham, C.C. 1954. An extension of the concept of partioning hereditary variance for the analysis of covariances among relatives when epistasis is present. Genetics 39:859-882.

Cockerham, C.C. 1956. Effects of linkage on the covariances between relatives. Genetics 41:138-141.

Cockerham, C.C. 1963. Estimation of genetic variances. Pages 53-94 *in* W.D. Hanson and H.F. Robinson, eds. Statistical genetics and plant breeding. Natl. Acad. Sci. - Natl. Res. Council Publ. 982.

Cockerham, C.C. 1971. Higher order probability functions of identity of alleles by descent. Genetics 69:235-246.

Cockerham, C.C. 1983. Covariances of relatives from self-fertilization. Crop Sci. 23:1177-1180.

Cockerham, C.C. 1984a. Covariances of relatives for quantitative characters with drift. Pages 195-208 *in* A. Chakravarti, ed. Human population genetics: The Pittsburgh Symposium. Van Nostrand Reinhold Co.

Cockerham, C.C. 1984b. Additive by additive variance with inbreeding and linkage. Genetics 108:487-500.

Cockerham, C.C., and D.F. Matzinger. 1985. Selection response based on selfed progenies. Crop Sci. 25:483-488.

Cockerham, C.C., and H. Tachida. 1988. Permanency of response to selection for quantitative characters in finite populations. Proc. Natl. Acad. Sci. 85:1563-1565.

Cockerham, C.C., and B.S. Weir. 1973. Descent measures for two loci with some applications. Theor. Pop. Biol. 4:300-330.

Cockerham, C.C., and B.S. Weir. 1977. Digenic descent measures for finite populations. Genet. Res. 30:121-147.

Cockerham, C.C., and B.S. Weir. 1984. Covariances of relatives stemming from a population undergoing mixed self and random mating. Biometrics 40:157-164.

Comstock, R.E., and H.F. Robinson. 1952. Estimation of average dominance of genes. Pages 495-516 *in* J.W. Gowen, ed. Heterosis. Iowa State College Press, Ames, Iowa.

Cooke, P., and K. Mather. 1962. Estimating the components of continuous variation. II. Genetical. Heredity 17:211-236.

Cooke, P., R.M. Jones, K. Mather, G.W. Bonsall, and J.A. Nelder. 1962. Estimating the components of continuous variation. I. Statistical. Heredity 17:115-133.

Cornelius, P.L., and J.W. Dudley. 1976. Genetic variance components and predicted response to selection under selfing and full-sib mating in a maize population. Crop Sci. 16:333-339.

Cornelius, P.L., and D.A. Van Sanford. 1988. Quadratic components of covariances of inbred relatives and their estimation in naturally self-pollinated species. Crop Sci. 28:1-7.

Cowen, N.M. 1986. Selection theory for selfed progenies. Theor. Appl. Genet. 73:182-189.

Croissant, G.L., and J.H. Torrie. 1971. Evidence of nonadditive effects and linkage in two hybrid populations of soybeans. Crop Sci. 11:675-677.

Crosby, J.L. 1973. Computer simulation in genetics. John Wiley and Sons, Toronto.

Crow, J.F., and M. Kimura. 1970. An introduction to population genetics theory. Harper & Row, New York.

Denniston, C. 1974. An extension of the probability approach to genetic relationships: one locus. Theor. Pop. Biol. 6:58-75.

Denniston, C. 1975. Probability and genetic relationship: two loci. Ann. Hum. Genet. 39:89-104.

Ewens, W.J. 1979. Mathematical population genetics. Springer-Verlag, Berlin.

Falconer, D.S. 1981. Introduction to quantitative genetics. 2nd ed. Longman, London.

Fisher, R.A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. Trans. Roy. Soc. Edinburgh 52:399-433.

Fisher, R.A., F.R. Immer, and O. Tedin. 1932. The genetical interpretation of statistics of the third degree in the study of quantitative inheritance. Genetics 17:107-124.

Fraser, A.S., and D. Burnell. 1970. Computer models in genetics. McGraw Hill, New York.

Frederickson, L.J., and W.E. Kronstad. 1985. A comparison of intermating and selfing following selection for heading date in two diverse winter wheat crosses. Crop Sci. 25:555-560.

Gallais, A. 1970. Covariances entre apparentes quelconques avec linkage et epistasie. I. Expression generale. Ann. Genet. Sel. Anim. 2:281-310.

Gallais, A. 1974. Covariances between arbitrary relatives with linkage and epistasis in the case of linkage disequilibrium. Biometrics 30:429-446.

Gates, C.E., R.E. Comstock, and H.F. Robinson. 1957. Generalized genetic variance and covariance formulae for self-fertilized crops assuming linkage. Genetics 42:749-763.

Gates, C.E., C.R. Weber, and T.W. Horner. 1960. A linkage study of quantitative characters in a soybean cross. Agron. J. 52:45-49.

Geldermann, H., and H. Gundel. 1979. Computer simulation. Pages 121-137 *in* J.N. Thompson,Jr. and J.M. Thoday, eds. Quantitative genetic variation. Academic Press, New York.

Gill, J.L. 1965. Effect of finite size on selection advance in simulated genetic populations. Aust. J. Biol. Sci. 18:599-617.

Gillois, M. 1964. La relation d'identite en genetique. These Fac. Sci. Universite de Paris, Paris. 295 pp.

Griffing, B. 1956. Concept of general and specific combining ability in

relation to diallel crossing system. Aust. J. Biol. Sci. 9:463-493.

Guttman, I. 1982. Linear models: An introduction. Wiley, New York.

Haldane, J.B.S. 1949. The association of characters as a result of inbreeding and linkage. Ann. Eugen. 15:15-23.

Hallauer, A.R., and J.B. Miranda, Fo. 1981. Quantitative genetics in maize breeding. Iowa State University Press, Ames, IA.

Hanson, W.D., and C.R. Weber. 1961. Resolution of genetic variability in self-pollinated species with an application to the soybean. Genetics 46:1425-1434.

Harris, D.L. 1964. Genotypic covariances between inbred relatives. Genetics 50:1319-1348.

Harville, D.A. 1977. Maximum likelihood approaches to variance component estimation and to related problems. J. Am. Stat. Ass. 72:320-338.

Hayman, B.I. 1954. The theory and analysis of diallel crosses. Genetics 39:789-809.

Hayman, B.I. 1955. The description and analysis of gene action and interaction. Cold Spring Harbor Symp. Quant. Biol. 20:79-86.

Hayman, B.I. 1960. Maximum likelihood estimation of genetic components of variation. Biometrics 16:369-381.

Hayman, B.I., and K. Mather. 1955. The description of genic interaction in continuous variation. Biometrics 11:69-82.

Horner, T.W., and C.R. Weber. 1956. Theoretical and experimental study of self-fertilized populations. Biometrics 12:404-414.

Horner, T.W., R.E. Comstock, and H.F. Robinson. 1955. Nonallelic gene interactions and the interpretation of quantitative genetic data. North Carolina Agr. Exp. Sta. Tech. Bull. 118. 117 pp.

Jacquard, A. 1974. The genetic structure of populations. (transl. from French.). Springer-Verlag, Berlin.

Jameson, D.L., ed. 1977. Evolutionary genetics. Dowden, Hutchinger & Ross, Inc., Stroudsburg, PA.

Jinks, J.L. 1954. The analysis of continuous variation in a diallel cross of *Nicotiana rustica* varieties. Genetics 39:767-788.

Jinks, J.L., and H.S. Pooni. 1982. Predicting the properties of pure breeding lines extractable from a cross in the presence of linkage. Heredity 49:265-270.

Jones, R.M. 1960. Linkage distributions and epistacy in quantitative inheritance. Heredity 15:153-159.

Kearsey, M.J., and J.L. Jinks. 1968. A general method of detecting additive, dominance and epistatic variation for metrical traits: I. Theory. Heredity 23:403-409.

Kempthorne, O. 1954. The correlation between relatives in a random mating population. Proc. Roy. Soc. London B143:103-113.

Kempthorne, O. 1955. The theoretical values of correlations between relatives in random mating populations. Genetics 40:153-167.

Kempthorne, O. 1956. On the covariances between relatives under selfing with general epistacy. Proc. Roy. Soc. London B145:100-108.

Kempthorne, O. 1957. An introduction to genetic statistics. Wiley, New York.

Kendall, M.G., and A. Stuart. 1963. The advanced theory of statistics. Vol. 1. Distribution theory. Hafner Pub. Co., New York.

Kennedy, Jr., W.J., and J.E. Gentle. 1980. Statistical computing. Marcel Dekker, Inc., New York.

Knott, D.R. 1987. The application of breeding procedures to wheat. Pages 419-427 *in* E.G. Heyne, ed. Wheat and wheat improvement. 2nd ed. ASA-CSSA-SSSA, Inc., Madison, WI.

Kojima, K., and R.C. Lewontin. 1970. Evolutionary significance of linkage and epistasis. Pages 367-388 *in* K. Kojima, ed. Mathematical topics in population genetics. Springer-Verlag, Berlin.

Li, C.C. 1976. First course in population genetics. Boxwood, Pacific Grove, CA.

Malecot, G. 1948. Les mathematiques de l'heredite. Masson, Paris.

Malecot, G. 1969. The mathematics of heredity. (Transl. from French.) Freeman, San Francisco, CA.

Mather, K. 1949. Biometrical genetics. Methuen, London.

Mather, K. 1979. Historical overview: quantitative variation and polygenic systems. Pages 5-34 *in* J.N. Thompson,Jr. and J.M. Thoday, eds. Quantitative genetic variation. Academic Press, New York.

Mather, K., and J.L. Jinks. 1971. Biometrical genetics. 2nd ed. Chapman and Hall, London.

Mather, K., and J.L. Jinks. 1982. Biometrical genetics. 3rd ed. Chapman and Hall, London.

Matzinger, D.F., and C.C. Cockerham. 1963. Simultaneous selfing and partial diallel test crossing. I. Estimation of genetic and environmental parameters. Crop Sci. 3:309-314.

Mayo, O. 1987. The theory of plant breeding. 2nd ed. Clarendon Press, Oxford.

Miller, P.A., and J.O. Rawlings. 1967. Breakup of initial linkage blocks through intermating in a cotton population. Crop Sci. 7:199-204.

Minvielle, F. 1987. Dominance is not necessary for heterosis: a two-locus model. Genet. Res. 49:245-247.

Moll, R.H., M.F. Lindsey, and H.F. Robinson. 1964. Estimates of genetic variances and level of dominance in maize. Genetics 49:411-423.

Mulitze, D.K. 1983. A critique of biometrical methods for estimating the number of genes controlling quantitative traits. Ph.D. Thesis. University of Saskatchewan, Saskatoon. 260 pp.

Narain, P. 1965. Homozygosity in a selfed population with an arbitrary number of linked loci. J. Genet. 59:254-266.

Nasoetion, A.H., C.C. Cockerham, and D.F. Matzinger. 1967. Simultaneous selfing and partial diallel test crossing. II. An evaluation of two methods of estimation of genetic and environmental variances. Biometrics 23:325-334.

Naylor, T.H., J.L. Balintfy, D.S. Burdick, and K. Chu. 1968. Computer simulation techniques. Wiley, New York.

Nelder, J.A. 1952. Some genotypic frequencies and variances occurring in biometrical genetics. Heredity 6:387-394.

Nelder, J.A. 1960. The estimation of variance components in certain types of experiment on quantitative genetics. Pages 139-158 *in* O. Kempthorne, ed. Biometrical genetics. Pergamon Press, New York.

Qureshi, A.W., and O. Kempthorne. 1968. The role of finite population size and linkage in response to continued truncation selection. Theor. Appl. Genet. 38:249-255.

Pederson, D.G. 1974. Arguments against intermating before selection in a self-fertilizing species. Theor. Appl. Genet. 45:157-162.

Pooni, H.S., and J.L. Jinks. 1982. Comparative analysis of association and dispersion crosses to detect linkage and epistatic components of variation. Heredity 49:211-220.

Pooni, H.S., and J.L. Jinks. 1986. Estimation of the true additive genetic variance $\sum_{i=1}^{k} d_i^2$ in the presence of linkage disequilibrium. Heredity 57:341-344.

Robertson, A. 1959. Experimental design in the evaluation of genetic parameters. Biometrics 15:219-226.

Robertson, A. 1970. A theory of limits in artifical selection with many linked loci. Pages 246-288 *in* K. Kojima, ed. Mathematical topics in population genetics. Springer-Verlag, Berlin.

Robertson, A. 1977. Artifical selection with a large number of linked loci. Pages 307-322 *in* E. Pollak, O. Kempthorne and T.B. Bailey, Jr., ed. Proc. Int. Conf. on Quantitative Genetics. Iowa State University Press, Ames, IA.

Robinson, F.H., and R.E. Comstock. 1955. Analysis of genetic variability in corn with reference to probable effects of selection. Cold Spring Harber Symp. Quant. Biol. 20:127-136.

Scheinberg, E. 1968. Methodology of computer genetic research. Can. J. Genet. Cytol. 10:754-761.

Schnell, F.W. 1961. Some general formulations of linkage effects in inbreeding. Genetics 46:947-957.

Schnell, F.W. 1963. The covariances between relatives in the presence of linkage. Pages 468-483 *in* W.D. Hanson and H.F. Robinson, eds. Statistical genetics and plant breeding. Natl. Acad. Sci. - Natl. Res. Council Publ. 982.

Sentz, J.C. 1971. Genetic variances in a synthetic variety of maize estimated by two mating designs. Crop Sci. 11:234-238.

Shikata, M. 1965. A generalization of the inbreeding coefficient. Biometrics 21:665-681.

Silvela, L., and R. Diez-Barra. 1985. Recurrent selection in autogamous species under forced random mating. Euphytica 34:817-832.

Strickberger, M.W. 1968. Genetics. The Macmillan Company, New York.

Thoday, J.M. 1961. Location of polygenes. Nature 191:368-370.

Thompson, Jr., J.N., and T.N. Kaiser. 1979. Computer simulation of the breeding program for polygene location. Pages 235-261 *in* J.N. Thompson Jr. and J.M. Thoday, eds. quantitative genetic variation. Academic Press, New York.

Turbo Pascal. 1986. Reference manual. Version 3.0. Borland International Inc., Scotts Valley, CA.

Van Aarde, I.M.R. 1975. The covariance of relatives derived from a random mating population. Theor. Pop. Biol. 8:166-183.

VandeLogt, M.B. 1978. Variance components and linkage effects in two barley crosses. M.S. Thesis. University of Wisconsin, Madison. 56 pp.

VandeLogt, M.B., M.A. Brinkman, and R.A. Forsberg. 1984. Variance components and linkage effects for quantitatively inherited traits in two barley crosses. Crop Sci. 24:323-327.

Van der Veen, J.H. 1959. Tests of non-allelic interaction and linkage for quantitative characters in generations derived from two diploid pure lines. Genetica 30:201-232.

Weber, W.E. 1982. Estimation of genetic variance components under linkage and sampling errors in self-fertilizing crops. Agronomie 2:201-212.

Weir, B.S., and C.C. Cockerham. 1969. Group inbreeding with two linked loci. Genetics 63:711-742.

Weir, B.S., and C.C. Cockerham. 1973. Mixed selfed and random mating at two loci. Genet. Res. 21:247-262.

Weir, B.S., and C.C. Cockerham. 1977. Two-locus theory in quantitative

genetics. Pages 247-269 *in* E. Pollak, O. Kempthorne and T.B. Bailey Jr., eds. Proc. Int. Conf. on Quantitative Genetics. Iowa State University Press, Ames, IA.

Weisberg, S. 1985. Applied linear regression. 2nd ed. Wiley, New York.

Wricke, G., and W.E. Weber. 1986. Quantitative genetics and selection in plant breeding. de Gruyter-Verlag, Berlin.

Wright, A.J. 1987. Covariance of inbred relatives with special reference to selfing. Genetics 115:545-552.

Wright, A.J. 1988. Some applications of the covariances of relatives with inbreeding. Pages 10-20 *in* B.S. Weir, E.J. Eisen, M.M. Goodman and G. Namkoong, eds. Proc. 2nd Int. Conf. on quantitative Genetics, Sinauer Assoc., Inc., Sunderland, MA.

Wright, A.J., and C.C. Cockerham. 1985. Selection with partial selfing. I. Mass selection. Genetics 109:585-597.

Wright, A.J., and C.C. Cockerham. 1986a. Selection with partial selfing. II. Family selection. Crop Sci. 26:261-268.

Wright, A.J., and C.C. Cockerham. 1986b. Covariances of relatives and selection response in generations of selfing from an outcrossed base population. Theor. Appl. Genet. 72:689-694.

Wright, S. 1921. Systems of mating. Genetics 6:111-173.

Wright, S. 1935. The analysis of variances and the correlations between relatives with respect to deviations from an optimum. J. Genet. 30:243-256.

Appendices

Appendix A: Program listing of GATESTEST.PAS

```
{*********************************************************}
Program GATESTEST_PAS;

{Programmed by Rong-Cai Yang and R.J. Baker}
{              University of Saskatchewan, 1988}

{*********************************************************}

Procedure COVS_PAS;

{This procedure is to generate genetic variances and covariances
 of relatives from diffferent generations of selfing under
 different genetic models and linkage values}

type realarray = array[1..1] of real;

var gf,gm,gn,gmw,gnw,temp,model,link,gnum : ^realarray;
    f,cov,lambda : real;
    i,j,l,u,v,w,k,kp,m,n,corr,md,total : integer;
    nr1,nc1,nr2,nc2,nr3,nc3 : integer;
    answer,response : char;
    fname : string[20];
    outfil : text;

{$i c:\pascal\matrix.inc}
{$i c:\yang\simula\function.inc}
{$i c:\yang\common.inc}

{These three commands indicate that three files from three different
 sub-directories of Drive C were included in this program:
 matrix.inc is a set of procedures which was developed by R.J. Baker
 and can be used to have matrix manipulation in this program;
 function.inc is a set of function which are not available
 Turbo PASCAL;
```

```
 Common.inc is a set of procedures commonly used in every major
 procedure of this program.}

Procedure genotypic_frequency(var mat; lambda : real;
                              k : integer; corr : integer);


var coef1,coef2,coef3,coef4 : real;
    a : ^realarray absolute mat;
    gf : array[1..10] of real;
    p,q,temp : real;
    i : integer;

begin
     p := (1.0 - lambda)/2.0;
     q := 1.0 - p;
     coef1 := power(0.5,k-1);
     coef2 := power(lambda/2.0,k-1);
     coef3 := power((0.5 - p*q),k-1);
     coef4 := 1.0/(1.0 + 2.0*p);
     gf[1] := 0.5*(coef4 - coef1 - 0.5*lambda*coef4*coef2 + 0.5*coef3);
     gf[2] := 0.5*(coef1 - coef3);
     gf[3] := p*coef4 - 0.5*(coef1 - 0.5*lambda*coef2*coef4
              - 0.5*coef3);
     gf[4] := gf[2];
     gf[5] := 0.5*(coef2 + coef3);
     gf[6] := 0.5*(-coef2 + coef3);
     if corr = -1 then begin
        temp := gf[3]; gf[3] := gf[1]; gf[1] := temp;
        temp := gf[6]; gf[6] := gf[5]; gf[5] := temp;
        end; {calculation of genotypic freq. for repulsion linkage}
     gf[7] := gf[4];
     gf[8] := gf[3];
     gf[9] := gf[2];
     gf[10] := gf[1];
     for i := 1 to 10 do
         a^[i] := gf[i];
end; {genotypic frequuency}

procedure gcoeff(var mat; lambda : real; t : integer; f : real);


var i,j,k : integer;
    c1,c2,c3,c4,p,q,x,y : real;
    a : ^realarray absolute mat;
    gw : array[1..10,1..8] of real;

procedure metric(k,i,j : integer);
```

```
begin
   gw[k,1] := 1.0*i - 1.0;
   x := 1.0 - sqr(gw[k,1]);
   gw[k,2] := 0.5*(f-1.0) + c1*x;
   gw[k,3] := 1.0*j - 1.0;
   y := 1.0 - sqr(gw[k,3]);
   gw[k,4] := 0.5*(f-1.0) + c1*y;
   gw[k,5] := gw[k,1]*gw[k,3] + c3*x*y;
   gw[k,6] := gw[k,1]*gw[k,4];
   gw[k,7] := gw[k,3]*gw[k,2];
   gw[k,8] := 0.25*sqr(f-1.0) + 0.5*(f-1.0)*c1*(x + y) + c4*x*y;
end; {metric}

begin {gcoeff}
   p := 0.5 - 0.5*lambda;
   q := 1.0 - p;
   c1 := power(0.5,t);
   c2 := power(0.5*lambda,t);
   c3 := (lambda/(2.0 - lambda))*(1.0 - c2);
   c4 := power((0.5 - p*q),t);

k := 0;
   for i := 2 downto 0 do
     for j := 2 downto 0 do
     begin
        k := k + 1;
        metric(k,i,j);
        if (i = j) and (i = 1)
           then
              begin
                k := k + 1;
                c3 := -c3;
                 metric(k,i,j);
              end; {then}
    end; {for i,j}
  for i := 1 to 10 do
        for j := 1 to 8 do
        begin
            a^[(i-1)*8+j] := gw[i,j];
        end; {for i,j}
end; {gcoeff}

function covar(var mat; k,m,n,corr : integer; lambda : real) : real;

var modl : ^realarray absolute mat;
    x,y,z : real;
    i : integer;
```

```
begin
   genotypic_frequency(gf,lambda,k,corr);
   gcoeff(gmw,lambda,m-k,f);
   mmult(gmw,modl,gm,10,8,1);
   gcoeff(gnw,lambda,n-k,f);
   mmult(gnw,modl,gn,10,8,1);

   x := 0.0; y := 0.0; z := 0.0;
   for i := 1 to 10 do
   begin
      x := x + gf^[i]*gm^[i];
      y := y + gf^[i]*gn^[i];
      z := z + gf^[i]*gm^[i]*gn^[i];
   end;
   covar := z - x*y;
end; {covar}

begin { main of COVS_PAS }
   clrscr;
   gotoxy(1,5);
   writeln(' Is F2-metric (y) or Fint-metric (n) used in this run? ');
   readln(answer);
   if answer in ['y','Y'] then f := 0.0;
   if answer in ['n','N'] then f := 1.0;
{Read the data files from the directory}
   writeln;writeln(' Read in the genetic models from the file... ');
   writeln;writeln(' How many rows and colomns of this file ? ');
   readln(nr1);readln(nc1);
   getmem(model,nr1*nc1*sizeof(real));
   control_read(model,nr1,nc1);
   writeln;writeln(' Read in generation numbers from the file... ');
   writeln;writeln(' How many rows and columns of this file ? ');
   readln(nr2);readln(nc2);
   getmem(gnum,nr2*nc2*sizeof(real));
   control_read(gnum,nr2,nc2);
   writeln;writeln(' Read in linkage values from the file... ');
   writeln;writeln(' How many rows and columns of this file ? ');
   readln(nr3);readln(nc3);
   getmem(link,nr3*nc3*sizeof(real));
   control_read(link,nr3,nc3);
{Reserve space for different matrices}
   getmem(gf,10*sizeof(real));
   getmem(gm,10*sizeof(real));
   getmem(gn,10*sizeof(real));
   getmem(gmw,10*8*sizeof(real));
   getmem(gnw,10*8*sizeof(real));
   getmem(temp,8*sizeof(real));
{Output the results from the simulation}
```

```
    writeln;writeln(' Enter the output file name for Cov[k,kp;m,n] ? ');
    writeln(' Suggested name of input file is GATESCOV.DAT');
    readln(fname);
    assign(outfil,fname);
    rewrite(outfil);
    writeln(outfil);writeln(outfil);

    total := 0;
    for l := 1 to nr3 do
    begin
      corr := trunc(link^[(l-1)*2+1]);
      lambda := link^[(l-1)*2+2];
      md := 0;
      for i := 1 to nr1 do
      begin
        md := md + 1;
        for j := 1 to nc1 do
        begin
          temp^[j] := model^[(i-1)*nc1+j];
        end;
        for u := 1 to nr2 do
        begin
          total := total + 1;
          k := trunc(gnum^[(u-1)*4+1]);
          kp := trunc(gnum^[(u-1)*4+2]);
          m := trunc(gnum^[(u-1)*4+3]);
          n := trunc(gnum^[(u-1)*4+4]);
          if kp = 1 then
             cov := covar(temp,k,m,n,corr,lambda)
             else cov := covar(temp,k,m,n,corr,lambda)
                         - covar(temp,kp,m,n,corr,lambda);
          writeln(outfil,k:2,kp:2,m:2,n:2,corr:3,
                              lambda:5:2,md:3,cov:8:4);
        end;
      end;
     writeln(' Count for total combinations:', total:10);
     end;
    close(outfil);
end; { main of COVS_PAS }

{*****************************************************************}

Procedure GATES_PAS;

{This procedure is to calculate the coefficient of determination
 from the reduced model in which only additive and dominance effects
 but no linkage are considered}
```

```
type realarray = array[1..1] of real;

var i,j,l,k,m,n,nparm,nvar,nh,loop,s,t,r,df,corr,md,ncv,nobs : integer;
    cch,ds,rs,f,resid,residh,cd,ccd,cda,ccda,cdh,cdha,ccdh,lk : real;
    z1,z2,z3,z4,z5,z7 : integer; z6 : real;
    hd,gcof,covp,covs,x,xh,y,b,bh,ssy,ssyh,ssr,ssrh : ^realarray;
    xp,yp,xpx,xpy,xpxi,bp : ^realarray;
    response : char;
    fname : string[20];
    infile,outfil : text;

{$i c:\pascal\matrix.inc}
{$i c:\yang\simula\function.inc}
{$i c:\yang\common.inc}

begin { main of GATES_PAS }
   clrscr;
   gotoxy(1,5);
   writeln;writeln(' How many parameters in full model ? ');
   readln(nparm);
   writeln;writeln(' How many parameters in the reduced model ?');
   readln(nh);
   writeln;writeln(' How many dependent variables ? ');
   readln(nvar);
   getmem(b,nparm*sizeof(real));
   getmem(bh,(nh)*sizeof(real));
   getmem(x,30*nparm*sizeof(real));
   getmem(xh,30*(nh)*sizeof(real));
   getmem(y,30*sizeof(real));
   getmem(gcof,30*6*sizeof(real));
   getmem(ssy,nvar*sizeof(real));
   getmem(ssyh,nvar*sizeof(real));
   getmem(ssr,nvar*sizeof(real));
   getmem(ssrh,nvar*sizeof(real));
   getmem(hd,ncv*6*sizeof(real));
   nobs := 30;
   getmem(xp,nparm*nobs*sizeof(real));
   getmem(yp,nvar*nobs*sizeof(real));
   getmem(xpx,nparm*nparm*sizeof(real));
   getmem(xpy,nparm*nvar*sizeof(real));
   getmem(xpxi,nparm*nparm*sizeof(real));
   getmem(bp,nvar*nparm*sizeof(real));

   writeln;writeln(' Give the output file for the ANOVA? ');
   readln(fname);
   assign(outfil,fname);
   rewrite(outfil);
   writeln(outfil);
```

```
    writeln;writeln(' Now read in Gates X1-X6....');
    control_read(gcof,30,6);
    gates(x,gcof,1);
    gates(xh,gcof,nh+1);
    assign(infile,'c:\yang\simula\gatescov.dat');
    reset(infile);
    for k := 1 to 300 do
    begin
       for i := 1 to 30 do
        readln(infile,z1,z2,z3,z4,z5,z6,z7,y^[i]);
        least_squares(x,y,ssy,ssr,b,30,1,nparm);
        least_squares(xh,y,ssyh,ssrh,bh,30,1,nh);
        cd := ssr^[1]/ssy^[1];
        ccd := 1.0 - cd;
        cdh := ssrh^[1]/ssy^[1];
        ccdh := 1.0 - cdh;
        ds := ssr^[1]-ssrh^[1]; {Difference between S.S.'s}
        rs := ssy^[1]-ssr^[1]; {Residual S.S.}
        f := (ds/4.0)/(rs/24.0);
        cch := cd-cdh;
        writeln(outfil,z5:3,z6:5:2,z7:3,cd:8:4,cdh:8:4,ssrh^[1]:10:3);
        writeln;
    end;
    close(infile);
    close(outfil);
end; { main of GATES_PAS }

{************************************************************}

Procedure SAMPLING_PAS;

{This procedure is to estimate the probability of rejecting the null
 hypothesis that there is no linkage under different models and linkage
 values with 1000 computer runs}


type realarray = array[1..1] of real;

var covs,gcoef,gcovs,pcovs,cf,sv : ^realarray;
    b,bh,x,xh,gcof,ssy,ssyh,ssr,ssrh : ^realarray;
    xp,yp,xpx,xpy,xpxi,bp : ^realarray;
    i,j,k,s,t,hs,ls,ncv,reps,nparm,nh,nvar,z,z5,counter : integer;
    cont,z6,cd,cdh,ds,rs,f : real;
    fname : string[20];
    outfil : text;
    ch : char;

{$i c:\pascal\matrix.inc}
```

```
{$i c:\yang\simula\function.inc}
{$i c:\yang\common.inc}

procedure randomadd(var mat1,mat2; m,n : integer; x : real);

var a : ^realarray absolute mat1;
    b : ^realarray absolute mat2;
    i,j : integer;
    mean,stdev : real;

begin
   mean := 0.0;
   for i := 1 to m do
   begin
     for j := 1 to n do
     begin
        stdev := x*a^[(i-1)*n+j];
        b^[(i-1)*n+j] := a^[(i-1)*n+j] + norm(mean,stdev);
     end;
   end;
end; {randomadd}

begin { main of SAMPLING_PAS }
   clrscr;
   gotoxy(1,5);

   writeln;writeln(' How many covs in the input file ? ');
   readln(ncv);
   writeln;writeln(' How many parameters in full model ? ');
   readln(nparm);
   writeln;writeln(' How many variables are dropped in this run ?');
   readln(nh);
   writeln;writeln(' How many dependent variables ? ');
   readln(nvar);

   getmem(b,nparm*sizeof(real));
   getmem(bh,(nh)*sizeof(real));
   getmem(x,30*nparm*sizeof(real));
   getmem(xh,30*nh*sizeof(real));
   getmem(gcof,30*6*sizeof(real));
   getmem(ssy,nvar*sizeof(real));
   getmem(ssyh,nvar*sizeof(real));
   getmem(ssr,nvar*sizeof(real));
   getmem(ssrh,nvar*sizeof(real));
   getmem(covs,ncv*8*sizeof(real));
   getmem(gcovs,30*sizeof(real));
   getmem(pcovs,30*sizeof(real));
   getmem(cf,4*sizeof(real));
```

```
getmem(sv,4*sizeof(real));
getmem(xp,nparm*30*sizeof(real));
getmem(yp,nvar*30*sizeof(real));
getmem(xpx,nparm*nparm*sizeof(real));
getmem(xpy,nparm*nvar*sizeof(real));
getmem(xpxi,nparm*nparm*sizeof(real));
getmem(bp,nvar*nparm*sizeof(real));

writeln;writeln(' Read in input file containing Covs ... ');
writeln;control_read(covs,ncv,8);
writeln;writeln(' Read in input file containing constants ... ');
writeln;control_read(cf,4,1);
writeln;writeln(' Read in sampel size, N = ');
read(reps);
writeln;writeln(' Give the output file for the F-values? ');
readln(fname);
assign(outfil,fname);
rewrite(outfil);
writeln(outfil);
writeln;writeln(' Now read in Gates X1-X6....');
control_read(gcof,30,6);
gates(x,gcof,1);
gates(xh,gcof,nh+1);
counter := 0;
for k := 1 to 2 do
begin
  for i := 1 to 30 do
  begin
     gcovs^[i] := covs^[((k-1)*30+i-1)*8+8];
  end;
  z5 := trunc(covs^[((k-1)*30+30-1)*8+5]);
  z6 := covs^[((k-1)*30+30-1)*8+6];
  for t := 1 to 4 do
  begin
    cont := cf^[t];
    ls := 0;
    for j := 1 to reps do
    begin
       randomadd(gcovs,pcovs,30,1,cont);
       least_squares(x,pcovs,ssy,ssr,b,30,1,nparm);
       least_squares(xh,pcovs,ssyh,ssrh,bh,30,1,nh);
       cd := ssr^[1]/ssy^[1];
       cdh := ssrh^[1]/ssy^[1];
       ds := ssr^[1]-ssrh^[1]; {Difference between S.S.'s}
       rs := ssy^[1]-ssr^[1]; {Residual S.S.}
       f := (ds/4.0)/(rs/24.0);
       if f > 2.78 then ls := ls + 1;
   end;
```

```
      counter := counter + 1;
      writeln(' Count for loops',counter:8);
      writeln(outfil,reps:5,cont:8:5,z5:3,z6:5:2,ls:5);
    end;
  end;
  close(outfil);
end; { main of SAMPLING_PAS }

begin { main of GATESTEST}
COVS_PAS;
GATES_PAS;
SAMPLING_PAS;
end. { main of GATESTEST}
```

Appendix B: Derivation of the relation $N > (1/\Delta)^2$

Let p be the probability of rejecting the null hypothesis of no linkage when

$$df_2(R_F^2 - R_R^2)/df_1(1 - R_F^2) > F_{\alpha\,(df_1,\,df_2)},$$

where R_F^2 is coefficient of determination for full model while R_R^2 is that for reduced model and df_1 and df_2 in Gates' test are 4 and 24, respectively. Then

$$p = \lim_{N \to \infty} \frac{n}{N},$$

where n is the number of times of rejecting H_0 and N is the total number of computer replicate runs. Thus, the estimate of this probability is

$$\hat{p} = \frac{n}{N}$$

with standard error

$$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}.$$

Choose N so that $|\hat{p} - p| < \Delta p$, but

$$|\hat{p} - p| < t_{\alpha,\,N-1} s_{\hat{p}} \simeq 2\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$$

Note that $\sqrt{\hat{p}(1-\hat{p})/N}$ is maximum at $p = 0.5$. If N is chosen so that

$$2\sqrt{\frac{0.5(1-0.5)}{N}} < \Delta p$$

then it will be sufficiently large for any p, where $0 \leq p \leq 1$. Thus,

$$N > \left(\frac{1}{\Delta p}\right)^2$$

Appendix C: Climatological data for Saskatoon

Year	May	June	July	August
		Monthly mean temperature (°C)		
1951-80	11.2	15.6	18.4	17.2
1986	12.7	15.7	17.1	17.2
1987	14.1	18.9	18.3	14.8
		Total precipitation (mm)		
1951-80	39.6	59.9	56.2	35.1
1986	56.0	52.7	115.7	27.7
1987	33.5	22.0	26.3	44.3
		Total evaporation (mm)		
1951-80	210.0	223.0	229.0	206.0
1986	221.6	240.0	185.1	251.0
1987	228.8	265.6	210.8	157.2

Source: Environment Canada data