A COMPUTER PROGRAM FOR SIMULATING GENETIC VARIANCES IN A "SELFING" POPULATION

R. R. Corbeil

Abstract

This program was written as an instructional and research aid to investigate the effects of genetic linkage, interaction and selection on heritable variance in populations of self-reproducing diploid organisms.

A COMPUTER PROGRAM FOR SIMULATING GENETIC VARIANCES IN A "SELFING" POPULATION

BU-407-M R. R. Corbeil January 1972

I. Introduction

This program was written as an instructional and research aid to investigate the effects of genetic linkage, interaction and selection on heritable variance in populations of self-reproducing diploid organisms. The computer model consists of ten loci arranged into two equal blocks with recombination between the loci under the control of three linkage parameters. Two of these parameters designate the intensity of linkage between loci in each of the two blocks separately and one specifies the linkage association between the blocks. This arrangement permits for the simulation of genetic systems consisting of from one to ten gene pairs and a diversity of linkage schemes.

The basic program sequence of instructions consists of evaluating the genetic constitution of a germ cell with respect to the ten loci mentioned above. Subsequently, two gametes are produced from this germ cell by a process involving the calculation of random deviates which are used by the program to select among all possible gametic configurations in random fashion. These are then fused and the resulting zygote scored for its numerical value. Each value thus generated is saved for statistical analysis and the corresponding zygotes are optionally stored to form the basis of a possible future generation.

The variation in zygotic values produced by this program is, therefore, of genetic origin. Gametic disequilibrium (or sampling error) is a necessary consequence of the process. No facility for simulating environmental error is provided.

A moderate number of input and output options are available to provide flexibility of problem applications. One such option is capable of producing a zygote

by zygote analysis, while another option will allow only summary statistics to be calculated and printed. Certain program limitations are related to capacity by the storage area dimension specified. These may be altered to suit individual needs or installation restrictions. A number of checks are built-in which may force the re-parameterization of a problem to fit existing program constraints. The model is implemented in the I.B.M. supported FORTRAN IV (E) language and may be readily modified to operate on any medium size computer system which supports FORTRAN logical statements. The programming techniques are essentially after the suggestions of A. Fraser and D. Burnell (1970).

II. Components of genetic variance

A. General background

Studies of multiple character inheritance in selfing populations can be traced directly to Mendel's work. He noted that for a particular locus only half of the progeny, derived from an initial hybrid, would exhibit further segregation. As a corollary Mendel also pointed out that the inheritance of non-discretely variable traits could be explained in terms of a multiplicity of segregating factors.

Shortly after the re-discovery of Mendel's work, Nilsson-Ehle in 1909 (see Provine [1971]) demonstrated the polygenic nature of kernel color in wheat. He observed that some lines yielded progeny with approximately 1/16 of the plants in each of the extreme color classes, whereas other lines approached the fraction 1/64. He explained these results as examples of two-factor and three-factor inheritance, respectively. He also reported data on a possible four-factor cross, and in conclusion indicated that a character controlled by as few as ten pairs of genes may be impossible to classify discretely; i.e., 3^{10} or approximately

60,000 phenotypes could be expected out of a possible 4^{10} genotypes.

With an increasing number of segregating units a continuum can quickly be established in the expression of the character. Under these circumstances genetic variation is further confounded by environmental sources of variation and the genetic studies of such traits becomes a statistical problem of partitioning the variance into its component sources. It wasn't until 1932 when R. A. Fisher, F. R. Immer, and O. Tedin published their treatise on genetic distributions that a method was formulated for extracting components of variation in selfing populations. K. Mather (1949), B. I. Hayman and K. Mather (1955) provided the main thrust in adapting Fisher's basic methods to breeding problems. These stressed the estimation of variance components. Hayman (1958) developed similar models based on generation means and L. Powers (e.g., see L. Powers, L. F. Locke and J. C. Garrett [1950]) provided an alternative to methods of component estimation that has yet to be fully explored.

Variation in phenotype may be viewed as changes manifest in an organism in the course of development. Many such processes are well known to be under genetic regulation, but it is rather more usual in polygenic inheritance to take variation to mean variance in the expression of a trait among individuals in a defined population (e.g., mature individuals of one filial generation). Certainly within the life span of one individual its genetic endowment remains relatively well fixed, but unlike coloration for example certain traits as stature are progressive and reflect different states of genetic activity at different times. In reality, unrecognized plysiological conditions (reflecting particular gene activities) will give rise to variations in measurement. An adequate experimental design will randomize such perturbations but should also be made to detect the major inequities arising from their interactions with the environment.

The most direct and obvious approach in partitioning phenotypic variance into its genetic and environmental components is to estimate the environmental source from known non-segregating populations. Once freed of environmental effects the residual variance may be considered to be due to genetic sources. Of these, assortment, segregation, epistasis, and linkage are paramount. Each is now considered briefly.

In the simplest feasible case, the additive-dominance model, variation of phenotypes within a generation is ascribed completely to segregation at each locus and their assortment. K. Mather and J. L. Jinks (1971) present this model in some detail. They partition genetic variance into that due to fixable variation (D), that is the additive or incremental effect of each locus summed over all homozygous loci; and the unfixable (H) in that it is due to the dominance effect at each locus summed over all heterozygous loci. The contribution of D and H to the expected variance of each generation is calculated from the premise that the genetic system behaves in a simple Mendelian fashion.

Only Mather's (1949) first few equations are presented here:

$$
\begin{aligned}
V_{F2} &= 1/2\ D + 1/4\ H \\
V_{\bar{F}\bar{3}} &= 1/2\ D + 1/16\ H \\
\bar{V}_{F3} &= 1/4\ D + 1/8\ H \\
W_{F23} &= 1/2\ D + 1/8\ H\ .
\end{aligned}
$$

All variances and the covariance (W_{F23}) are directly estimable from the experimental data. If each F_2 individual is allowed to produce a family by selfing then the variance of the means of such families is an estimate of $V_{\bar{F}\bar{3}}$ once an appropriate reduction due to environmental error has been made. The variance of

each F3 family can also be calculated and their average is an estimate of $\bar{V}_{F3}$ again after an appropriate reduction due to error. The covariance, which is without an environmental error term, is estimated from each F3 family mean and its F2 parent value.

The four variance expectations derived from an F2 population and F3 families may be extended by similarly considering F4 families. The number of equations possible increases geometrically with the increase in number of generations pursued. In the additive-dominance model, however, two equations are sufficient to estimate D and H. The value of the additional equations provided by F3 families is in the information which may be gained from them. Tests for detecting linkage and epistasis have been formulated by Mather based on the expected heterogeneous contribution of each to D and H in different generations. In addition, calculating the variance of the F3 within family variances (V_{VF3}) yields a statistic in the form of $1/4\ kx^2$, where k is the number of loci (see Mather and Jinks [1971] p. 311). Since $\bar{V}_{F3} = 1/2\ kx$ in the equivalent form then $(\bar{V}_{F3})^2/V_{VF3} = k$, which is a minimum estimate of the number of loci operating in the genetic system under study.

As the number of equations increase with the number of generations produced more elaborate models can be built. D. Robson (see K. Daly and D. Robson [1969]) has developed equations in which all possible digenic type interactions may be estimated from variances. Up to eleven parameters are necessary to characterize backcross generations. Only four of nine interaction terms are involved in V_{F2}, $V_{\underline{\quad}}$, and $\bar{V}_{F3}$ in addition to the familiar D and H terms.
$\overline{F3}$

Mather and Jinks (1971) investigate epistasis by its disturbances in D and H. In the presence of epistasis the expected V_{F2} becomes

$$1/2 \sum\left(d_a + 1/2 \sum_{ja.}\right) + 1/4 \sum\left(h_a + 1/2 \sum_{\ell a.}\right) + 1/4\ i^2 + i/8\ j^2 + 1/16\ \ell^2$$

where D is now estimated by $\sum(d + i/2\Sigma j)^2$ and H by $\sum(h + 1/2\Sigma\ell)^2$; i^2 is the sum of the additive loci by additive loci interactions j^2 the additive-dominance interactions and ℓ^2 the dominance-dominance interactions.

The interaction components and their influence on D and H may be summarized as:

$$V_{F2} = 1/2 \sum\left(d_a + 1/2 \sum_{ja\cdot}\right)^2 + 1/4 \sum\left(h_a + 1/2 \sum_{\ell a\cdot}\right) + 1/2\, i^2 + 1/8\, j^2 + 1/16\, \ell^2$$

$$V_{\underline{\quad}F3} = 1/2 \sum\left(d_a + 1/4 \sum_{ja\cdot}\right)^2 + 1/16 \sum\left(h_a + 1/4 \sum_{\ell a\cdot}\right) + 1/4\, i^2 + 1/32\, j^2 + 1/256\, \ell^2$$

$$\bar{V}_{F3} = 1/4 \sum\left(d_a + 1/4 \sum_{ja\cdot}\right)^2 + 1/8 \sum\left(h_a + 1/4 \sum_{\ell a\cdot}\right) + 5/16\, i^2 + 7/64\, j^2 + 1/32\, \ell^2$$

and $$W_{F23} = 1/2 \sum\left(d_a + 1/2 \sum_{ja\cdot}\right)\left(d_a + 1/4 \sum_{ja\cdot}\right) + 1/8 \sum\left(h_a + 1/2 \sum_{\ell a\cdot}\right)\left(h_a + 1/4 \sum_{\ell a\cdot}\right)$$
$$+ 1/4\, i^2 + i/16\, j^2 + 1/64\, \ell^2,$$

where a = 1,2,3,...,k for k loci.

The effects of linkage appear less readily detectable than those of epistasis, especially if recombination between pairs of loci is greater than 35% or less than 15%. Otherwise, linkage manifests itself in the estimates of D and H which is notably expressed as a departure of the ratio $V_{F2}/2\ \bar{V}_{F3}$ from unity. Possible interpretations of a rise or fall in the ratio is discussed by Mather (1949).

The simulation of genetic systems of linkage, di-, tri-, and polygenic interactions applied singly or in combinations, offers a clear avenue for investigating the extent of such disturbances on the usefulness of present analytical procedures for field application.

B. Programming considerations

The simulation of variances usually implies the creation of a population by digital computer whose elements are derived singly under a stochastic process so as to have the collection reflect an underlying distribution with only chance deviations. Translated into a genetic system of naturally selfing organisms this means simulating segregation and assortment, the consequences of meiosis. The F1 hybrid upon selfing could be expected to give progeny whose genotypic distribution in the F2 would be binomial with parameters $n = 20$ and $\theta = 0.5$. In succeeding generations only n would change by a factor of 1/2 each generation. However, with intra-locus interaction (dominance), and inter-locus interactions (epistasis) and linkage not only possible but usually present in some combination the binomial model becomes naïve.

In order to avoid the need to formulate a general mathematical model and still retain program flexibility the simulation process was made to reflect the biological mechanisms directly. Success in achieving behavioral similarity between biological phenomenon and the computer model rests upon an adequate understanding of the biological rather than the mathematical. Certain advantages accrue from this approach: it more closely approximates the genetic mechanisms as they are understood, it allows for an easy genetic re-parameterization between problems with little concern for expressing underlying distributions, and it has greater heuristic value for the geneticists and students it is intended to serve. The program procedures utilized are outlined in the last paragraph of the introductory section. They are expanded below in parallel to program flow.

The first genetic parameters evaluated in the program after an initialization and status check procedure are the recombination values R_1, R_2, and R_3 which have been assigned by the user. Each is expressed in the continuum from zero to 0.5 in conformity with genetic convention.

The arrangement of the ten loci into blocks of five each allows for more efficient computer use, reducing the amount of programming needed and the size of storage to be allocated. The motivation for blocking however, arose from several field observations of possible gene clusters (see R. R. Corbeil and L. Butler [1964]). Schematically, there are 2^5 possible gametic genotypes for each block and the probability that any one will be included as part of a gamete is a function of R_i. For a complete gamete this may be expressed as $1/2 \cdot R_1^r(1-R_1)^{4-r}R_3^s(1-R_3)^{1-s}R_2^t(1-R_2)^{4-t}$ where r, s, and t are the number of cross-overs in a random walk along the chromatids of the zygote yielding a particular gamete; r and t can take any integer value from 0 through 4 but s = 0 or 1.

The cumulative distributions (c.d.) corresponding to R_1 and R_2 are calculated separately, each yielding 32 class intervals. A random deviate from the uniform distribution with parameters (0,1) is generated to provide a random access into one of these c.d. intervals. The index of each interval also identifies a unique recombination mask. The procedure is repeated for each block and finally, a third random deviate is compared with R_3 to determine whether a crossover is to be effected between blocks.

The 32 recombination masks are the binary coded integers 0 to 31, i.e., one unit less than the index of its corresponding c.d. interval. Once randomly selected the recombination mask is logically manipulated along with the parent zygote to yield a gamete. Say that the zygote has the following genotype:

$\underline{+\ +\ +\ +\ +}$, and that the mask selected is in the form + - - - + (NOTE: This

$-\ -\ -\ -\ -$

mask represents a double crossover), then

$$+\ +\ +\ +\ +\ \text{⊗}\ +\ -\ -\ -\ +\ \rightarrow\ +\ -\ -\ -\ +\qquad (1)$$

$$-\ -\ -\ -\ -\ \text{⊗}\ -\ +\ +\ +\ -\ \rightarrow\ -\ -\ -\ -\ -\qquad (2)\ .$$

The logical sum of (1) and (2) → + - - - + (3). In (1) the mask is logically multiplied to one chromatid and in (2) the complement of the mask is multiplied to the other chromatid and the two results are summed. The logical entities + and -, representing contrasting alleles, are substituted by T and F, respectively, in the program. (See Ch. 1 of Fraser and Burnell [1970] for a thorough exposition.) The second half of the ten locus gamete is determined in similar fashion. If a recombination between blocks is indicated this tantamounts to incorporating the complement of the second block of genes derived in the formation of the gamete rather than its original derivative. Finally, the entire gametic forming program sequence is repeated and the two resulting gametes are fused into a single progeny zygote.

Scoring the zygote genotypically, locus by locus, is accomplished in two steps: i) the number of homozygous standard (+ +), heterozygous (+ - or - +), and homozygous mutant (- -) loci for the zygote in question is assessed, and ii) the numerical value of the zygote is calculated as the sum of the individual loci contributions and all pairwise (digenic) interactions. The five genetic parameters necessary in this evaluation are provided by the user. Up to 1000 such zygotic scores may be saved for statistical analysis. The zygotes themselves may be stored upon option to form the basis of a future generation.

A program facility exists for displaying each zygote generated along with its hexadecimal representation and numerical score. Display of various intermediate statistics may also be opted. Please refer to the section on program procedures and input options for details (Sec. III. B).

There are three levels at which summary statistics are calculated. At the initial level, which usually is made to simulate an F2 generation, the mean and variance of the total collection of zygotes is computed. Subsequent generations arise from "selfed" individuals and, therefore, distinct families are produced.

Hence, the intermediate level involves the computation of the mean and the variance of each family. These intermediate statistics form the collection from which third level summary statistics are derived. These are: the variance of means (e.g., $V_{\overline{F3}}$), the mean variance (e.g., $\bar{V}_{F3}$), the variance of the family variances (e.g., V_{VF3}), and the covariance of the family mean and its parent value (e.g., W_{F23}). Estimates of D, H, the number of genes which segregated, and the average number of loci remaining heterozygous are included also.

In order to keep the program's size in the vicinity of 60K of core requirement a total limit of 1000 zygotes per level is dimensioned in the FORTRAN source program. In the F3 generation a limit of 1000 families is imposed and in this limit only one offspring per family can be saved to form the basis of the F4 generation. This limitation imposes, therefore, constraints on the number of variances capable of being simulated in the F4 and thereafter. In fact, though each subsequent generation has the potential of introducing a new level for statistical summarization no new level is introduced. The family statistics for generation F_n are each from an individual of generation F_{n-1} but no two individuals from one F_{n-1} family can yield F_n progeny if the number of F_{n-1} families is 1000. Where the number of F_{n-1} families is less than 1000 it is possible to have two or more individuals from each family producing F_n progeny. No effort, however, is made to collect these into groups.

For selfing populations in the absence of selection relatively few generations are needed before all loci become fixed in the homozygous state. The program, therefore, is designed to terminate the processing of an active problem once genetic variance is reduced to near zero by passing control to the next problem in the queue.

As alluded to above the program is operable in the selection mode and great flexibility exists in applying different selection criteria within a single

problem phase. Selection is applied to the numerical value of the zygote supplied by the previous generation. The expected range of zygotic expression can be divided by the user into as many as three classes with each being assigned a different parameter of fitness. The parameters assigned by the user are automatically normalized thus optimizing the search for acceptable zygotes. The selection procedure places a zygote into one of the available classes. The decision to accept or reject this zygote is made by comparing the fitness value of its class with a random deviate on the uniform (0,1) distribution. If the random deviate is larger than its fitness value the zygote is rejected. The selection process continues until a user defined limit becomes saturated or the pool of available zygotes is exhausted.

This procedure of selection even if directed in favor of intermediate values does not long delay the process of genetic fixation in general. However, selecting for overdominance or selection under other special genetic schemes can delay fixation and even establish a balanced polymorphism. Only user imagination is likely to limit the number of genetic schemes which can be explored by this program.

The routine used for generating all random deviates was constructed after the suggestions of Prof. George E. Forsythe of the Computer Science Department, Stanford University. The routine is repeated several times rather than made into a subroutine because this has proven to be far more efficient. Several tests on the distribution of the random deviates and on the results from actual simulations have yielded highly satisfactory results. For a more extensive discussion of pseudo-random generators and methods of testing their output the reader is referred to Tocher (1963).

III. Procedure

A. Introduction

An effort was made to provide simplicity and consistency in the operation of the program through its variety of options. As far as was feasible input data fields were kept uniform and blank fields were left to designate default options. A small number of data checks have been provided in anticipation of detecting what are likely to be common input errors and incompatible execution-time developments. Some of these checks call for messages to be displayed, and all are designed to prevent abnormal job termination.

Program data input and output control is under the FORTRAN commands READ and PRINT without reference to data sets. Certain options call for hexadecimal input or output which may not be supported in certain installations; the pertinent instructions, however, may easily be modified (see appendix). Several instances of mixed-mode arithmetic may also be readily converted where the need exists. The appendix holds further information pertinent to implementation and the FORTRAN IV source listing.

B. Input

All input data is in the form of cards, and with one exception may be classed as either i) program control or ii) genetic parameters. The exception is the very first card of the data deck which initializes the pseudo-random number generators:

Cols. 1-9 IX: is any integer number consisting of one to nine digits right-justified in the field. It initializes the routines which generate random deviates. It is output on the printer for reference. It should be randomly chosen for each job. It must appear first and only once in a data deck.

1. Program control cards.

Each program control card corresponds to one biological generation. It specifies the program options desired, family size, maximum population size, mode of selection and the several selection indices.

These program variables are listed as follows:

Col. 1 FLAG: takes an integer value 0 to 9. 0 - - signals the end of a job. A blank card placed last in the data deck will suffice.

1 - - calls for input of genetic parameters. Whenever a FLAG = 1 option is stated the card immediately following in the data deck should be a genetic parameters card (see below for a description of these program variables). This is an obligatory set-up option, usually it produces an F2 from a hybrid zygote. Consequently, the first biological generation to be produced, whether subsequent generations will be computed or not, should characteristically be identified by the FLAG = 1 option.

2 - - this option requires a prior execution of an option FLAG = 1. It characteristically identifies all subsequent biological generations. It also forces the computation of individual family statistics.

3 through 9 - - may be used to signal the end of a current problem.

Col. 2 SAVZYG: takes an integer value of 0 through 4. 0 - - under this option, the zygotes produced during the course of program execution are not saved or stored and, therefore, are not available to future generations. The output of intermediate statistical summaries associated with each

family of an F3, F4, etc. is inhibited. 1 - - is similar to option SAVZYG = 0 with the exception that the available statistical summaries are printed.

* 2 - - under this option a pool of zygotes is saved so that they may serve as the basis of a future generation. the intermediate statistical summaries are not printed.

* This option is routinely the most useful and practical.

3 - - differs from SAVZYG = 2 in that the intermediate statistical summaries are printed.

4 - - activates all available data output options. Each zygote produced is saved and displayed through the printer along with its hexadecimal code and its numerical value. In addition all statistical summaries, by family and by generation, are printed.

Cols. 3-6 NZW: takes an integer value of 0 through 1000 and must be right-justified in its field. This program variable specifies the number of zygotes wanted (i.e., to be retained) from the zygotic pool of the previous generation. Hence, NZW also determines the number of families which are to be produced in the current generation. NZW automatically takes a value of zero when FLAG = 1 option is in effect.

Cols. 7-10 NPOP: takes an integer value of 0 through 1000 and must be right-justified in its field. This variable specifies the size of the initial generation under a FLAG = 1 option and the size of each family under a FLAG = 2 option.

Cols. 11-14 are left blank.

Col. 15 KC: takes an integer value of 0 through 9. It specifies the selection mode of program operation.

0 - - specifies operation is under no selection, may be left blank.

1 through 9 - - coupled with specification of ASL, BSL, WLO, WMID, and WHI below indicates operation under selection mode.

Cols. 16-20 ASL: is a positive or negative real variable. Omission of a decimal point in its field will be interpreted as XXX.XX, where the X's represent columns 16 through 20. This number specifies an internal boundary within the range of the zygotes' genotypic values at which fitness changes. This value is specified only under the KC≠0 options. ASL = A - selection level.

Cols. 21-25 BSL: a real variable similar to ASL above. It specifies a second internal boundary at which fitness changes. BSL = B - selection level.

Cols. 26-30 WLO: is a real variable taking any value from 0.0 to 1.0. Unless a decimal appears in the field the entry is regarded as .XXXXX WLO is not specified when option KC = 0. It represents the relative fitness or Malthusian parameter associated with genotypes whose numerical value are smaller than the ASL value.

Cols. 31-35 WMID: is similar to program variable WLO but corresponds to genotypic values falling on or between ASL and BSL.

Cols. 36-40 WHI: is similar to WLO and corresponds to genotypic values which are greater than BSL.

2. The genetic parameters

Each problem must have at least one and usually only one card containing the genetic parameters. These include three recombination values, five linear components for calculating genotypic values, and optionally the genotype of the founding hybrid in hexadecimal code.

Cols. 1-5 R1: takes real number values from 0.0 through 0.5 which represents the probability with which a recombination (cross-over) can occur between any two contiguous genes in the first block of five loci. Omission of the decimal point in this field is interpreted as .XXXXX and blanks are interpreted as zeros.

Cols. 6-10 R2: is similar to program variable R1 except that it specifies the probability of cross-over in the second block of five loci.

Cols. 11-15 R3: is similar to R1 but specifies the probability of cross-over between the two blocks.

Cols. 16-20 ED: is a real program variable taken in the form XXX.XX if the decimal point is omitted from its field. It specifies the increment contributed to the genotypic value of a zygote by each homozygous standard locus. It also specifies the decrement for each homozygous mutant locus.

Cols. 21-25 EH: is of the same numerical type as ED above. It specifies the contribution, plus or minus, of each heterozygous locus to the genotypic value.

Cols. 26-30 EDD: is also of the same numerical type as ED above. It is an interaction component between all pairs of homozygous loci. The contribution to the genotypic value follows:

$$\left[\frac{ND \times (ND-1)}{2} \times EDD \right] + \left[\frac{NR \times (NR-1)}{2} \times EDD \right] +$$

$[ND \times NR \times (-EDD)]$, where ND = number of homozygous standard loci in the zygote and NR = number of homozygous mutant loci.

Cols. 31-35 EDH: is also of the same numerical type as ED above. It is the interaction component due to all pairs of loci where one of the pair is homozygous and the other heterozygous. Its contribution is: $[ND \times NH \times EDH] + [NH \times NR \times (-EDH)]$, where NH = number of heterozygous loci in the zygote.

Cols. 36-40 EHH: is also of the same numerical type as ED above. It is the interaction component between all pairs of heterozygous loci. Its contribution is: $\frac{NH \times (NH-1)}{2} \times EHH$.

Cols. 41-48 A1A2B1B2: is optional and may be left blank in which case the hexadecimal value 1F1F0000 is assumed. This is equivalent to a zygote whose genotype is $\frac{+\ +\ +\ +\ +\ +\ +\ +\ +\ +}{-\ -\ -\ -\ -\ -\ -\ -\ -\ -}$, because A1 = 1F = 11111 (in binary) or + + + + + (in genetic symbolism) and, A2 = 1F also, B1 = 00 = 00000 (in binary) or - - - - - (in genetics), and B2 = 00 also. A1 and A2 represent the two blocks of genes in the upper chromatid while B1 and B2 represent the two blocks of corresponding alleles in the lower of the two homologous chromosomes.

Any hexadecimal value may be assigned to A1, A2, B1, B2 which is in the range of 00 to 1F. For example, an 08 assignment for A1 would represent a - + - - - gene arrangement and a 0815170B assignment for A1 through B2, respectively, would represent the zygote $\frac{A1A2}{B1B2} = \frac{- + - - - + - + - +}{+ - + + + - - + + +}$.

3. An example of a data deck

Card No.	Column No. 1
1	006354013
2	1200000500
3	.5^^^.5^^^.5^^^2.0^^1.0^^0.5^^0.5^^0.5^^
4	2000500025
5	1200000050
6	0.5^^0.5^^0.5^^2.0^^1.0^^0.0^^0.5^^0.0^^
7	2200500010^^^^1-20.^+20.^1.0^^0.0^^1.0^^
8	2200500010^^^^1-20.^+20.^1.0^^0.0^^1.0^^
9	2200500010^^^^1-20.^+20.^1.0^^0.0^^1.0^^
10	2000500010^^^^1-20.^+20.^1.0^^0.0^^1.0^^
11	blank card
NOTE:	^ represents a blank column

Explanation: Two problems are queued in this sequence of cards, the first involving cards 2 through 4 and the second 5 through 10.

Card no. 1 sets the variable IX equal to 6,354,013 which may be chosen from a random numbers table and initializes the pseudo-random number generators. This card only appears once in a job-step.

The first problem is one which simulates an experiment in which the polygenic system has 10 independent genes with all possible pairwise inter-loci type interactions. It is requested on card no. 2 that the F2 populations consist of 500 zygotes and on card no. 4 that the F3 consist of 50 families of 25 zygotes each.

The second problem is initiated in a similar fashion but specifies an F2 of only 50 zygotes. Cards nos. 7 through 10 represent generations 3 through 6 under a regime of selection which favors extreme genotypic values. Col. 15 specifies selection and cols. 16-40 contain the information pertinent to the desired selection scheme. Under the genetic parameters specified, the range of possible zygotic expressions is - 42.5 to + 42.5 because of the interaction. Without interaction the expected range would be - 20.0 to + 20.0 but zygotes bounded by these values are rejected 100% of the time, whereas, zygotes with values less than - 20 or greater than + 20 are preserved. On the chance that this population is not quickly extinguished selection would favor those zygotes with a homozygous-heterozygous loci interaction. Only with the maintenance of a number of heterozygous loci can this "selfing" population survive.

The results of a single trial run of the above problems are given in Section IV along with other sample results.

C. Output

Output options are largely under the control of the program variable SAVZYG discussed in the previous section. Further, but brief considerations are given here.

Mention should be made about the amount of output which can be generated, especially that of option SAVZYG = 4. In an F2 population with a size of 500 zygotes each zygote would be displayed in a form similar to that diagrammed in Sec. B.2 above. On many line printers this would represent approximately 25 pages of output at 60 lines/page. Subsequently an F3 specified as consisting of 50 families of 25 zygotes each, approximately another 65 pages of output would be generated. However, SAVZYG options may be used rather more judiciously than this as they may be changed within a problem in that they need to be re-specified for each generation.

The option SAVZYG = 3 can generate moderate amounts of output, output which is unnecessary in many cases. Its usefulness is to identify the distribution with which F3 families are produced under particular genetic schemes. In an F3 population consisting of 50 families approximately 3 printed pages of output would be generated.

Certain types of output cannot be suppressed. These are the F2 summary statistics and the summaries of subsequent generations operating under FLAG = 2 option.

A partial example of output under SAVZYG = 4 is given in Table 1.

TABLE 1

EXAMPLES OF OUTPUT

```
LINE REF.
   NO.

    1     012452199
    2     NO.        CHROMATID X/ CHROMATID Y      HEX CODED          VALUE
    3        1         T T T F T T T F F F         1D180F16           18.5000
    4                  F T T T T T F T T F

    5        2         T F F T T T F T T T         13171B06           13.5000
    6                  T T F T T T F F T T F

    7        3         T T F F T F T F T F         190A1E1F           26.0000
    8                  T T T T F T T T T T

    9     CASE 1       R1 = 0.5000   R2 = 0.5000   R3 = 0.5000

                       ED = 2.0   EH = 1.0   EDD = 0.0   EDH = 0.5   EHH = 0.0
   10                  NO. ZYGOTES = 5   MEAN = 11.6000   VAR = 132.05004883
   11     NO. 2        PARENT GENOTYPE = 190A1E1F     VALUE = 26.0000

   12     NO.        CHROMATID X/CHROMATID Y       HEX CODED          VALUE
   13        2         T T F F F F T F T F         180A181E           2.0000
   14                  T T F F F T T T T F

   15        5         T T F T T T T T T F         1B1E181A           13.50000
   16                  T T F F F T T F T F

   17     CASE 1       R1 = 0.5000   R2 = 0.5000   R3 = 0.5000

                       ED = 2.0   EH = 1.0   EDD = 0 0   EDH = 0.5   EHH = 0 0
   18                  SEL.CODE(KC) = 1   A-LEVEL  -20.000   B-LEVEL = 20.000
                       WLO = 1.000        WMID = 0.500       WHI = 1.000
   19                  NO. ZYGOTES = 2    MEAN = 7 7500      VAR. = 66 12500000

   20     SUMMARY STATISTICS FOR GENERATION NO. 3
   21     CASE 1       R1 = 0.500   R2 = 0.5000   R3 = 0.5000

                       ED = 2.0   EH = 1.0   EDD = 0.0   EDH = 0.5   EHH = 0.0

   22                  ESTIMATE OF NUMBER OF GENES SEGREGATING IN THE
                       PREVIOUS GENERATION = -0.6   THIS GENERATION = 1.7

   23                  AV. F3 = 6.250000   VALUES OF PARENTS USED--AV  = 14.6667
                       AND THEIR VAR. = 186.583374
   24                  VAR. F3 = 78.250000     MVAR F3 = 55.12500
                       COV. F3 = 191.750000    VV. F3 = 903.000000
   25                  SEL.CODE(KC) = 1   A-LEVEL = -20.000   B-LEVEL = 20.000
                       WLO = 1.000        WMID = 0.500       WHI = 1.000
   26                  NO. OF ZYGOTES SAVED (NZW) = 3     FAMILY SIZE (NPOP) = 2
                       D = 61.666656                      H = 317.666504
```

Under option SAVZYG = 0 only lines 1, 9, and 10 would have been printed in this example. Under SAVZYG = 1, lines 1, 9, 10, 11, and 17 through 26 would be printed, lines 11, 17-19 repeated for each family and lines 20-26 repeated for each subsequent generation. Option SAVZYG = 4 is similar to option SAVZYG = 1 in type of output with the addition of lines in the form of 3, and 4 for each zygote produced. As mentioned earlier, however, SAVZYG = 1 does not create a pool of zygotes from which a future generation can be developed.

IV. Results of sample problems

Nine problems were chosen to check program behavior. These were chosen because the expected values for many of the calculations performed could readily be computed from existing theory. The results are presented in Table 2.

A cursory examination of the tabulated statistics reveals good agreement between the calculated values and the simulation results. It should be noted that the $V_{\overline{F3}}$ may be improved in some cases by reducing the observed variance by the estimate of its sampling error, $1/25\ \bar{V}_{F3}$. Estimates of sampling error are not available for the other statistics.

The population and family sizes assigned to these sample problems are modest by simulation standards. In terms of plant breeding experiments, however, the 1,750 plots simulated here does not represent a trivial field experiment. Yet, one should not be inclined to consider dealing with lesser numbers in reality especially since additional facilities will be necessary to estimate environmental sources of variation.

TABLE 2

RESULTS OF NINE SAMPLE TRIALS

PARAMETERS					$\bar{F}_2$		$\bar{F}_3$		V_{F2}		$V_{\bar{F}\bar{3}}$		$\bar{V}_{F3}$		W_{F23}	
ED	EH	EDD	EDH	EHH	EXP.	OBS.	EXP.	OBS.	EXP.	OBS.	EXP.	OBS.	EXP.	OBS.	EXP.	OBS.
2	0	0	0	0	0.00	-.14	0.00	0.18	20.00	18.90	20.00	24.87	10.00	10.20	20.00	24.19
2	0	.5	0	0	0.00	-.26	0.00	0.01	22.81	21.77	22.81	19.91	13.52	13.27	22.81	19.80
2	0	0	.5	0	0.00	0.02	0.00	-.60	91.72	98.20	49.18	56.60	25.64	33.35	67.11	80.31
2	0	0	0	.5	5.63	5.54	1.41	0.88	33.36	33.50	20.83	14.67	11.93	11.15	23.34	16.91
2	1	0	0	0	5.00	5.18	2.50	2.38	22.50	21.17	20.63	21.23	11.25	10.99	21.25	22.22
2	1	.5	0	0	5.00	5.31	2.50	2.23	25.31	26.69	23.44	25.94	14.77	14.91	24.06	27.04
2	1	0	.5	0	5.00	4.87	2.50	0.70	94.22	85.70	49.80	40.98	26.89	25.93	68.36	56.59
2	1	0	0	.5	10.63	10.31	3.91	3.65	47.11	46.45	22.87	20.47	16.00	16.60	28.81	20.63
2	1	.5	.5	.5	10.63	10.58	3.91	3.78	121.64	122.14	54.86	63.49	35.16	34.45	78.73	91.31

$R_1 = R_2 = R_3 = 0.5$; F_2 = 500 zygotes; F_3 = 50 families; FAMILY = 25 zygotes

The two problems of the sample deck in Sec. III. B.3 were processed. The first of these is summarized by the last line of Table 2. The second provides an example of the program operating in the selection mode. A portion of the results generated is summarized in Table 3.

TABLE 3

A SELECTION EXPERIMENT

GENERATION	MEAN	VAR.	NZW		AVE.NO. LOCI SEG.		PARENTS	
NO.			SPECIFIED	OBSERVED	ESTIMATED	ACTUAL	MEAN	VAR.
2	6.01	94.22	50	50	11.7	10.0	10.00	0.00
3	18.97	2.91	50	3	-7.2	3.0	25.67	4.08
4	18.21	3.54	50	11	5.9	1.7	26.23	8.32
5	18.72	9.51	50	28	5.8	1.2	24.14	9.35
6	20.60	3.01	50	50	-	1.0	24.65	4.31

$R_1 = R_2 = R_3 = 0.5$, ED = 2.0, EH = 1.0, EDD = 0.0, EDH = 0.5, EHH = 0.0

FAMILY SIZE = 10, ASL = -20., BSL = 20., WLO = 1.0, WMID = 0.0, WHI = 1.0

Of the original 50 F2 zygotes in the selection experiment only three were selected to give rise to the F3 generation. As each produced ten progeny a pool of 30 zygotes was available to found the F4. Only eleven, however, were adequate vis-à-vis the selection criteria. By the sixth filial generation the population had returned to its prescribed level. On the average, each zygote in the F6 had retained one heterozygous locus crucial for expressing a zygotic value exceeding $\pm$ 20. A zygote with a single segregating locus has to be either homozygous standard at all other loci or homozygous mutant at all other loci to fall outside the range $\pm$ 20. Whether a segregating locus can be maintained in this simulation experiment is speculative, but with a family size of ten the chances are good and should average 0.5 in frequency. Each homozygous gene

produced at the locus in question has now effectively become lethal. The semblance of a balanced polymorphic system has evolved. Interesting questions of such a system evolving in simulation might be the incidence of success and failure, the consequence of spontaneous mutation, and the maintenance of heterozygosity in naturally inbreeding populations.

V. References

Corbeil, R. R. and L. Butler, 1964. The Role of Chromosome 2 in the Genetics of Maturation Time in the Tomato. Can. J. Genet. Cytol. 6:446-452.

Daly, K. and D. Robson, 1969. Estimates of Genetic Parameters from a Hybrid Derivative in Nicotiana. Genetics 62:201-213.

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. and D. Burnell, 1970. Computer Models in Genetics. McGraw-Hill Book Co., New York.

Hayman, B. I., 1958. The Separation of Epistatic from Additive and Dominance Variation in Generation Means. Hered. 12:371-390.

Hayman, B. I. and K. Mather, 1955. The Description of Genic Interactions in Continuous Variation. Biometrics 11:69-82.

Marsaglia, G. and T. A. Bray, 1968. One-Line Random Number Generators and Their Use in Combinations. Comm. ACM 11:757-759.

Mather, K. 1949. Biometrical Genetics. Methuen and Co., Ltd., London.

Mather, K. and J. L. Jinks, 1971. Biometrical Genetics, 2nd. ed. Cornell University Press, Ithaca, New York.

Powers, L., L. F. Locke and J. C. Garrett, 1950. Partioning Method of Genetic Analysis Applied to Quantitative Characters of Tomato Crosses. U.S.D.A. Tech. Bull. No. 998.

Provine, W. B. 1971. The Origins of Theoretical Population Genetics. The University of Chicago Press, Chicago.

Tocher, K. D. 1963. The Art of Simulation. The English Universities Press Ltd., London.

VI. Appendix

The program was written in I.B.M. supported FORTRAN IV. Installations supporting A.S.A. Standard FORTRAN IV will be capable of compiling this program once the following have been considered:

1. Support for hexadecimal input and output.
2. Support for mixed-mode arithmetic.

The statement number 35 in the source listing contains the specification 4Z2 which may readily be changed to 4I2 with a corresponding change of notation type to be used in the input field. For example, designation of a complete hybrid in this system under hexadecimal notation is 1F1F0000, whereas, in integer notation it becomes 31310000. The heading specified by statement no. 118 should also be modified from HEX to DEC for completeness The specification 4Z2 in statement no. 82 and Z10 in statement no. 204 should be changed to 4I2 and I10, respectively. Under the I-type specification blanks appear in the output field where leading zeros might be expected.

The statements containing mixed-mode arithmetic include: nos. 177, 217, 218, 220, 244-251, 255, and 256. All the mixed-mode arithmetic statements but no. 177 involve division by the integer variables NPOP or NZW.

The three statement sequences (statement nos. 61-63, 121-123, 135-137, 140-142, and 145-147) for generating pseudo-random numbers are designed to yield optimum results with hardware-specified 32 bit words. Other configurations may necessitate amending these built-in routines for improved simulation results (see G. Marsaglia and T. A. Bray [1968] for suggestions). Compilation under I.B.M. FORTRAN(H) is not advised without special notice being given to the distribution characteristics of the random deviates.

The compiled and link-edited program can operate under FORTRAN(G) in 64K of core. Two rather inefficient in-line routines have been incorporated into the program to reduce the data storage allotment by a factor of four. Under Cornell University's I.B.M. System 360/65 up to 150 zygotes/second can be produced, depending slightly on the type of problem. Each of the sample trials presented in Table 2 cost less than ¢0.50 for complete processing indicating the program is moderately efficient in operation.

2/72 aw

SOURCE LISTING

STATEMENT NUMBER

```
      C     A COMPUTER MODEL FOR SIMULATING GENETIC VARIANCES--
    1       DIMENSION FR(16),TF(2,32),R(3),A(2),B(2),AL(5),BL(5),
           1CHM(5),GAM1(2,5),GAM2(2,5),IZYG(10),ZYGOTE(1000),NZYG(
           C  1000),
           2VALUE(1000),AV(1000),VAR(1000),NZTBS(1000),RZYG(1000)
    2       DOUBLE PRECISION RC,CR,EXP1,EXP2,EXP3,EXP4,EXP5,FR,TF
    3       LOGICAL AL,BL,CHM,GAM1,GAM2
    4       INTEGER AJ,BJ,A,B,SAVZYG,FLAG
    5       IPASS = 0
      C     INITIALIZATION OF RANDOM NUMBER GENERATING ROUTINES
    6       READ 101, IX
    7     101 FORMAT (I9)
    8       PRINT 105, IX
    9     105 FORMAT (1H ,I9)
      C     DATA INPUT AND PROGRAM CONTROL
   10     100 NZW=0
   11       NRP=0
   12       1 IF (NZW.GT.0) GO TO 8
   13       2 NZF=0
   14         NSEG = 0
   15         NRP=NRP+1
   16         PRINT 103
   17     103 FORMAT (1H ,119H----------------------------------------
           C  ------------
           1-----------------------------------------------------
           C  ------------
           2---/)
   18         READ 301, FLAG,SAVZYG,NZW,NPOP,KC,ASL,BSL,WLO,WMID,WHI
   19     301 FORMAT (2I1,2I4,4X,I1,2F5.2,3F5.5)
   20         IF ((FLAG.EQ.2).AND.(IPASS.EQ.1)) GO TO 2
   21         IF (FLAG.EQ.1) NZW = 0
   22         IF (NZW.EQ.0) NG = 2
   23         IF (FLAG.EQ.0) CALL EXIT
   24         IF (FLAG.GE.3) GO TO 100
   25         IF ((NZW.GT.1000).OR.(NPOP.GT.1000)) GO TO 998
   26         IF ((SAVZYG.EQ.0).AND.(FLAG.EQ.1)) SAVZYG = 1
   27         IF (KC.EQ.0) GO TO 5
      C     NORMALIZATION OF MALTHUSIAN PARAMETERS
   28         ANORM = AMAX1 (WLO,WMID,WHI)
   29         ANORM = 1.0 / ANORM
   30         WLO = WLO * ANORM
   31         WMID = WMID * ANORM
   32         WHI = WHI * ANORM
   33       5 IF (FLAG.GT.1) GO TO 8
   34         READ 302, R(1),R(2),R(3),ED,EH,EDD,EDH,EHH,DA1,DA2,DB1
           C  ,DB2
   35     302 FORMAT (3F5.5,5F5.2,4Z2)
   36         IF ((DA1.EQ.0).AND.(DA2.EQ.0).AND.(DB1.EQ.0).AND.(DB2.
           C  EQ.0))
           1GO TO 6
   37         A(1) = DA1
   38         A(2) = DA2
   39         B(1) = DB1
   40         B(2) = DB2
```

```
41            GO TO 10
42          6 A(1) = 31
43            A(2) = 31
44            B(1) = 0
45            B(2) = 0
46            GO TO 10
      C       DATA INITIALIZATION FOR THE SECOND + NTH GENERATION
47          8 IF (NZF.EQ.NZW) GO TO 80
48            NZF=NZF+1
49            IF (NZF .NE. 1) GO TO 9
50            NRP=NRP-1
51            NG = NG + 1
      C       SELECTION ROUTINE ON GENOTYPIC VALUE
52            NZSEL = 0
53            NZSCR = 0
54        501 NZSCR = NZSCR + 1
55            IF ((NZSCR.GT.NSIZE).OR.(NZSCR.GT.1000)) GO TO 505
56            Q = RZYG(NZSCR)
57            IF (KC.EQ.0) GO TO 504
58            IF (Q.LT.ASL) W=WLO
59            IF ((Q.GE.ASL).AND.(Q.LE.BSL)) W=WMID
60            IF (Q.GT.BSL) W=WHI
61            IX = IX*314159269 + 453806245
62            IF (IX.LT.0) IX = IX + 2147483647 + 1
63            YFL = FLOAT(IX)*.4656613E-9
64            IF (YFL.GT.W) GO TO 501
65        504 NZSEL = NZSEL + 1
66            NZTBS(NZSEL) = NZYG(NZSCR)
67            VALUE(NZSEL) = Q
68            IF ((NZSEL.LT.NZW).AND.(NZSCR.LT.NSIZE).AND.(NZSCR.LT.
           C  1000))
           1GO TO 501
69        505 NZW = NZSEL
70            IF (NZW.LT.2) PRINT 106
71        106 FORMAT (1H ,6X,33HABOVE POPULATION HAS GONE EXTINCT)
72            IF (NZW.LT.2) GO TO 2
73          9 INT=NZTBS(NZF)
      C       EXPANDING THE FOUR BYTES OF HEXADECIMAL CODED GENOTYPE
      C          INTO
      C       FOUR FULL WORDS OF BINARY
74            A(1)=INT/16777216
75            INT=INT-A(1)*16777216
76            A(2)=INT/65536
77            INT=INT-A(2)*65536
78            B(1)=INT/256
79            B(2)=INT-B(1)*256
80            IF ((SAVZYG.EQ.0).OR.(SAVZYG.EQ.2)) GO TO 15
81            PRINT 102, NZF,A(1),A(2),B(1),B(2),VALUE(NZF)
82        102 FORMAT (1HO,6X,3HNO.,I4,3X,18HPARENT GENOTYPE = ,4Z2,5
           C  X,8HVALUE =
           1,F10.4)
83            GO TO 15
      C       CALCULATE AND STORE TEST FREQUENCIES
84         10 DO 13 J=1,2
```

```
      RC=R(J)
      CR=1.-RC
      EXP1=CR**4/2.
      EXP2=RC*CR**3/2.
      EXP3=RC*RC*CR*CR/2.
      EXP4=RC**3*CR/2.
      EXP5=RC**4/2.
      FR(1)=EXP1
      FR(2)=EXP2
      FR(3)=EXP3
      FR(4)=EXP2
      FR(5)=EXP3
      FR(6)=EXP4
      FR(7)=EXP3
      FR(8)=EXP2
      FR(9)=EXP3
      FR(10)=EXP4
      FR(11)=EXP5
      FR(12)=EXP4
      FR(13)=EXP3
      FR(14)=EXP4
      FR(15)=EXP3
      FR(16)=EXP2
      DO 11 K=1,32
   11 TF(J,K)=0.0
      TF(J,1)=FR(1)
      DO 12 K=1,15
   12 TF(J,K+1)=TF(J,K) + FR(K+1)
      DO 13 K=16,31
      KK = 32-K
   13 TF(J,K+1)=TF(J,K)+FR(KK)
C     DETERMINE GAMETE CONFIGURATIONS, ZYGOTE
   15 IF (SAVZYG.NE.4) GO TO 17
      PRINT 204
  204 FORMAT (1HO,30X,3HNO.,5X,23HCHROMATID X/CHROMATID Y,4X
     C  ,9HHEX CODED
     1,7X,5HVALUE/)
   17 DO 50 LP=1,NPOP
      DO 30 J=1,2
      IX = IX*314159269 + 453806245
      IF (IX.LT.0) IX = IX + 2147483647 + 1
      YFL = FLOAT(IX)*.4656613E-9
      I=0
   18 I=I+1
      IF ((YFL.GE.TF(J,I)).AND.(I.LT.32)) GO TO 18
      MASK = I - 1
      CALL DECODE (MASK,CHM)
      AJ=A(1)
      CALL DECODE (AJ,AL)
      BJ=B(1)
      CALL DECODE (BJ,BL)
      DO 19 L=1,5
   19 GAM1(J,L)=(AL(L).AND.CHM(L)).OR.(BL(L).AND..NOT.CHM(L)
     C  )
```

```
135           IX = IX*314159269 + 453806245
136           IF (IX.LT.0) IX = IX + 2147483647 + 1
137           YFL = FLOAT(IX)*.4656613E-9
138           IF (YFL.GE.R(3)) GO TO 23
139           IF (CHM(5).AND..TRUE.) GO TO 22
140        21 IX = IX*314159269 + 453806245
141           IF (IX.LT.0) IX = IX + 2147483647 + 1
142           YFL = FLOAT(IX)*.4656613E-9
143           YFL=YFL/2.
144           GO TO 24
145        22 IX = IX*314159269 + 453806245
146           IF (IX.LT.0) IX = IX + 2147483647 + 1
147           YFL = FLOAT(IX)*.4656613E-9
148           YFL=YFL/2.+.5
149           GO TO 24
150        23 IF (CHM(5).AND..TRUE.) GO TO 21
151           GO TO 22
152        24 I=0
153        25 I=I+1
154           IF ((YFL.GE.TF(J,I)).AND.(I.LT.32)) GO TO 25
155           MASK = I - 1
156           CALL DECODE (MASK,CHM)
157           AJ=A(2)
158           CALL DECODE (AJ,AL)
159           BJ=B(2)
160           CALL DECODE (BJ,BL)
161           DO 29 L=1,5
162        29 GAM2(J,L)=(AL(L).AND.CHM(L)).OR.(BL(L).AND..NOT.CHM(L)
             C  )
163        30 CONTINUE
      C       GENERATE ZYGOTE VECTOR, VALUE
164           DO 35 J=1,5
165           IZYG(J)=0
166           IZYG(J+5)=0
167           DO 35 I=1,2
168           IF (GAM1(I,J).AND..TRUE.) IZYG(J)=IZYG(J)+1
169        35 IF (GAM2(I,J).AND..TRUE.) IZYG(J+5)=IZYG(J+5)+1
170           NR=0
171           NH=0
172           ND=0
173           DO 40 I=1,10
174           IF (IZYG(I).EQ.0) NR=NR+1
175           IF (IZYG(I).EQ.1) NH=NH+1
176        40 IF (IZYG(I).EQ.2) ND=ND+1
177           SCORE=ND*ED+((ND*(ND-1))/2)*EDD+NH*EH+((NH*(NH-1))/2)*
             C  EHH+NR*(-ED)
             1+((NR*(NR-1))/2)*EDD+ND*NR*(-EDD)+ND*NH*EDH+NH*NR*(-ED
             C  H)
178           ZYGOTE(LP)=SCORE
179           NSEG = NSEG + NH
180           IF (NZF.NE.0) GO TO 45
181           LOOP = LP
182           IF (SAVZYG.GT.1) RZYG(LP) = SCORE
183           GO TO 48
```

```
184          45 LOOP = NZW*LP-NZW+NZF
185             IF((SAVZYG.GT.1).AND.(LP*NZW.LE.1000)) RZYG(LOOP) = SC
               C  ORE
186          48 IF ((SAVZYG.LE.1).OR.(LP*NZW.GT.1000)) GO TO 50
187             NZYG(LOCP)=0
188             DO 49 I=1,2
189             IF (GAM1(I,1).AND..TRUE.) NZYG(LOOP)=NZYG(LOOP)+16
190             IF (GAM1(I,2).AND..TRUE.) NZYG(LOOP)=NZYG(LOOP)+8
191             IF (GAM1(I,3).AND..TRUE.) NZYG(LOOP)=NZYG(LOOP)+4
192             IF (GAM1(I,4).AND..TRUE.) NZYG(LOOP)=NZYG(LOOP)+2
193             IF (GAM1(I,5).AND..TRUE.) NZYG(LOOP)=NZYG(LOOP)+1
194             NZYG(LOOP)=NZYG(LOOP)*256
195             IF (GAM2(I,1).AND..TRUE.) NZYG(LOOP)=NZYG(LOOP)+16
196             IF (GAM2(I,2).AND..TRUE.) NZYG(LOOP)=NZYG(LOOP)+8
197             IF (GAM2(I,3).AND..TRUE.) NZYG(LOOP)=NZYG(LOOP)+4
198             IF (GAM2(I,4).AND..TRUE.) NZYG(LOOP)=NZYG(LOOP)+2
199             IF (GAM2(I,5).AND..TRUE.) NZYG(LOOP)=NZYG(LOOP)+1
200             IF (I.LT.2) NZYG(LOOP)=NZYG(LOOP)*256
201          49 CONTINUE
202             IF (SAVZYG.NE.4) GO TO 50
203             PRINT 201, LOOP,(GAM1(1,I),I=1,5),(GAM2(1,J),J=1,5),NZ
               C  YG(LOCP),
               1SCORE
204         201 FORMAT (1H ,29X,I5,5X,10L2,5X,Z10,5X,F10.4)
205             PRINT 202, (GAM1(2,I),I=1,5),(GAM2(2,J),J=1,5)
206         202 FORMAT (40X,10L2/)
207          50 CONTINUE
C               SECOND GENERATION AND INTERMEDIATE SUMMARY STATISTICS
208             IPASS = 0
209             SUM=0.0
210             SS=0.0
211             DO 60 I=1,NPOP
212             SUM=SUM+ZYGOTE(I)
213          60 SS=ZYGOTE(I)*ZYGOTE(I)+SS
214             NZL = NZF
215             IF (NZF.EQ.0) NZL = 1
216             NSIZE = NPOP * NZL
217             ASEG = NSEG / NPOP
218             AV(NZL) = SUM / NPOP
219             VAR(NZL) = 0.0
220             IF (NPOP.GT.1) VAR(NZL) = (SS-(SUM*SUM/NPOP))/(NPOP-1)
221             IF (((SAVZYG.EQ.0).AND.(NZF.GT.0)).OR.((SAVZYG.EQ.2).A
               C  ND.(NZF.GT.0
               1))) GO TO 1
222             PRINT 205, NRP,R(1),R(2),R(3),ED,EH,EDD,EDH,EHH
223         205 FORMAT (1H ,6X,5HCASE ,I2,3X,4HR1 =,F8.4,2X,4HR2 =,F8.
               C  4,2X,4HR3 =,
               1F8.4,3X,4HED =,F5.1,2X,4HEH =,F5.1,2X,5HEDD =,F5.1,2X,
               C  5HEDH =,F5.1
               2,2X,5HEHH =,F5.1)
224             IF (KC.GT.0) PRINT 207, KC,ASL,BSL,WLO,WMID,WHI
225         207 FORMAT (1H ,16X,16HSEL. CODE (KC) =,I2,3X,9HA-LEVEL =,
               C  F7.3,3X,
               19HB-LEVEL =,F7.3,3X,5HWLO =,F6.3,3X,6HWMID =,F6.3,3X,5
```

```
     C  HWHI =,F6.3)
226        PRINT 203, NPOP, AV(NZL), VAR(NZL)
227   203 FORMAT (1H ,16X,13HNO. ZYGOTES =,I5,5X,6HMEAN =,F10.4,
     C  5X,6HVAR. =,
     1F20.8/)
228        GO TO 1
      C     SUMMARY STATISTICS FOR THE THIRD AND SUBSEQUENT GENERA
      C       TIONS
229       80 AVAL = 0.0
230        PVAR = 0.0
231        AVX3 = 0.0
232        V1X3 = 0.0
233        V2X3 = 0.0
234        W1X23 = 0.0
235        VVX3 = 0.0
236        DO 85 ILP = 1,NZW
237        V2X3 = VAR(ILP) + V2X3
238        VVX3 = VAR(ILP) * VAR(ILP) + VVX3
239        AVX3 = AV(ILP) + AVX3
240        AVAL = VALUE(ILP) + AVAL
241        PVAR = VALUE(ILP) * VALUE(ILP) + PVAR
242        W1X23 = AV(ILP) * VALUE(ILP) + W1X23
243       85 V1X3 = AV(ILP) * AV(ILP) + V1X3
244        W1X23 = (W1X23 - ((AVX3 * AVAL)/NZW))/(NZW-2)
245        VVX3 = (VVX3 - ((V2X3 * V2X3)/NZW))/(NZW-1)
246        V1X3 = (V1X3 - ((AVX3 * AVX3)/NZW))/(NZW-1)
247        PVAR = (PVAR - ((AVAL * AVAL)/NZW))/(NZW-1)
248        V2X3 = V2X3 / NZW
249        AVX3 = AVX3 / NZW
250        AVAL = AVAL / NZW
251        IF (NG.EQ.3) VF3M = V1X3 - (V2X3/NPOP)
252        D = 8.*(VF3M - (2.**(NG-4)*V2X3))/ 3.
253        H = 16. * (2.**(NG-2) * V2X3 - VF3M)/3.
254        GK = 0.0
255        IF ((VVX3.NE.0.0).AND.(NPOP.GT.1)) GK = V2X3*V2X3/(VVX
     C  3-(2.*V2X3*
     1V2X3/(NPOP-1)))
256        ESEG = ASEG / NZW
257        PRINT 210, NG
258   210 FORMAT (1H ,6X,37HSUMMARY STATISTICS FOR GENERATION NO
     C  .,I3)
259        PRINT 205, NRP,R(1),R(2),R(3),ED,EH,EDD,EDH,EHH
260        PRINT 217, GK, ESEG
261   217 FORMAT (1H ,16X,68HESTIMATE OF NUMBER OF GENES SEGREGA
     C  TING IN THE
     1PREVIOUS GENERATION =,F5.1,4X,17HTHIS GENERATION =,F5.
     C  1)
262        PRINT 211, NG,AVX3,AVAL,PVAR
263   211 FORMAT (1H ,16X,4HAV.F,I2,2H =,F10.4,9X,29HVALUES OF P
     C  ARENTS USED-
     1-AV. =,F10.4,5X,16HAND THEIR VAR. =,F12.6)
264        PRINT 212, NG,V1X3,NG,V2X3,NG,W1X23,NG,VVX3
265   212 FORMAT (1H ,16X,5HVAR.F,I2,2H =,F12.6,5X,6HMVAR.F,I2,2
     C  H =,F12.6,5X
```

```
     1,5HCOV.F,I2,2H =,F12.6,5X,4HVV.F,I2,2H =,F12.6)
266       IF (KC.GT.0) PRINT 207, KC,ASL,BSL,WLO,WMID,WHI
267       PRINT 216, NZW,NPOP,D,H
268   216 FORMAT (1H ,16X,28HNO. OF ZYGOTES SAVED (NZW) =,I4,4X,
     C  20HFAMILY SI
     1ZE (NPOP) =,I4,5X,3HD =,F12.6,4X,3HH =,F12.6/)
269       IF (VVX3.EQ.0.0) IPASS = 1
270       GO TO 2
271   998 PRINT 215, NRP
272   215 FORMAT (1HO,96HTHE CONTROL PARAMETERS DESIGNATED EXCEE
     C  D THE CAPACI
     1TY OF THE STORAGE AREA, EXECUTION OF CASE NO.,I2,15H I
     C  S TERMINATED
     2.)
273       READ 302, DUMMY
274       IPASS = 1
275       GO TO 2
276       END
```

```
277           SUBROUTINE DECODE (ICH,CHX)
278           DIMENSION CHX(5)
279           LOGICAL CHX
280           DO 5 IN=1,5
281         5 CHX(IN)=.FALSE.
282           IF (ICH.GE.16) CHX(1)=.TRUE.
283           IF (ICH.GE.16) ICH=ICH-16
284           IF (ICH.GE.8) CHX(2)=.TRUE.
285           IF (ICH.GE.8) ICH=ICH-8
286           IF (ICH.GE.4) CHX(3)=.TRUE.
287           IF (ICH.GE.4) ICH=ICH-4
288           IF (ICH.GE.2) CHX(4)=.TRUE.
289           IF (ICH.GE.2) ICH=ICH-2
290           IF (ICH.EQ.1) CHX(5)=.TRUE.
291           RETURN
292           END
```