Predicting and observing the properties of second cycle hybrids using basic generations and inbred line × F_1 crosses

**H. S. Pooni, J. L. Jinks and
J. F. F. de Toledo**

**Department of Genetics, University of Birmingham,
P.O. Box 363, Birmingham B15 2TT.**

The phenotypic and genotypic properties of second cycle hybrids are predicted from the means and variances of the F_2 generation of two crosses between pairs of inbred varieties of *Nicotiana rustica* ($V_1 \times V_5$ and $V_2 \times V_{12}$). The predicted properties are compared with the observed properties of the second cycle hybrids generated by crossing the first cycle recombinant inbred lines with the F_1 of the original cross (the L_3 generation of an inbred line triple test cross) which, in the absence of a linkage disequilibrium, are identical with the second cycle hybrids produced as a diallel set of crosses among the first cycle inbred lines. In the presence of a linkage disequilibrium the diallel set of crosses has a smaller linkage bias than the inbred line × F_1 crosses.

The agreement between the predicted and observed phenotypic properties of the second cycle hybrids is good over a wide range of characters in both crosses. The overall correlations between the observed and expected proportions are 0·81 and 0·88 for the $V_1 \times V_5$ and $V_2 \times V_{12}$ crosses, respectively. Furthermore, the genotypic predictions show that only few of the second cycle hybrids and second cycle inbred lines will be superior to the best first cycle inbred lines which were themselves superior to their own better parents and their heterotic F_1's.

Genetical explanations of the origins of the superior derivatives of an initial cross are explored using a development of the standard notation of Mather and Jinks (1982).

INTRODUCTION

In our previous paper (Toledo, Pooni and Jinks, 1984) we showed how to predict the phenotypic distribution of second cycle hybrids and to use it to predict the proportion of hybrids scoring higher or lower than the parents (P_1 and P_2) and the F_1 of the original cross (first cycle hybrid) or the best first cycle recombinant inbreds derived from this cross. In this paper we shall analyse an extensive set of data to illustrate the validity of the theory for predicting the performance of the hybrids in a wide range of genetical situations. We shall also introduce a notation for the second cycle hybrids, adapted from Mather and Jinks (1982) to avoid ambiguity in discussing the causes of the superior performance of such hybrids.

Finally predictions of the breeding value of a cross will be made in respect of its potential for producing superior first cycle recombinant inbreds and superior second cycle hybrids so that the relative advantages of these alternative end products of the breeding programme can be compared and related to the underlying genetical mechanisms.

The properties of the second cycle hybrids will be predicted from the mean and variance of the F_2 generation of the original cross while those of the first cycle inbreds will be predicted from the basic generations (P_1, P_2, F_1, F_2, B_1 and B_2) of the same cross. Hence the generations required for making these predictions will, in most cases, have already been raised for other purposes during the early stages of a breeding programme.

NOTATION

The mean of a second cycle hybrid genotype, derived from a cross where P_1 and P_2 differ at k loci assuming only additive (d) and dominance (h) genetic effects, can be defined as:

$$\bar{G} = m + \sum^{k_d} d - 2 \sum^{k'_d} d + \sum^{k-k_d} h,$$

where k_d is the total number of homozygous loci, k'_d is the number of these loci homozygous for the negative allele, hence ($\sum^{k_d} d - 2 \sum^{k'_d} d$) is the net balance of the additive genetic effects of the k_d

homozygous loci and $\sum^{k-k_d} h$ is the sum of the dominance effects at the remaining $(k-k_d)$ heterozygous loci.

In the presence of non-allelic interactions this expectation becomes

$$\bar{G} = m + \sum^{k_d} d - 2\sum^{k'_d} d + \sum^{k-k_d} h + \sum^{(1/2)k_d(k_d-1)} i$$
$$- 2\sum^{k'_d(k_d-k'_d)} i$$
$$+ \sum^{k_d(k-k_d)} j - 2\sum^{k'_d(k-k_d)} j + \sum^{(1/2)(k-k_d)(k-k_d-1)} l$$

where m is the mean of F_∞ inbred lines as defined by Van der Veen (1959); $\sum^{(1/2)k_d(k_d-1)} i$ is the summation of additive × additive interactions over all pairs of homozygous loci; $\sum^{k'_d(k_d-k'_d)} i$ is the summation of additive × additive interactions for the homozygous pairs of loci for which the alleles are in dispersion; $\sum^{k_d(k-k_d)} j$ is the summation of additive × dominance interaction for all pairs of homozygous/heterozygous loci; $\sum^{k'_d(k-k_d)} j$ is the summation of additive × dominance interactions for all pairs of homozygous/heterozygous loci where decreasing allele is fixed at the homozygous locus of the pair and $\sum^{1/2(k-k_d)(k-k_d-1)} l$ is the summation of the dominance × dominance interactions between all pairs of heterozygous loci. Further the additive × additive (i), additive × dominance (j) and dominance × dominance (l) interactions of each pair of loci can take positive or negative values based on the type of epistasis present.

In the conventional notation of Mather and Jinks (1982) the definitions of G and G' for any genotype "ii" reduce to

$$\bar{G}_{ii} = m + [d]_{ii} + [h]_{ii} \quad \text{and}$$

$$\bar{G}'_{ii} = m + [d]_{ii} + [h]_{ii} + [i]_{ii} + [j]_{ii} + [l]_{ii}$$

where $[d]_{ii}$, $[h]_{ii}$, $[i]_{ii}$, $[j]_{ii}$ and $[l]_{ii}$ represent the net contributions of additive, dominance, additive × additive, additive × dominance and dominance × dominance genetic effects after internal cancellations. Each component is potentially unique for each genotype and it will vary in magnitude and direction according to the proportions of homozygous (k_d) to heterozygous $(k-k_d)$ loci and the degree of gene dispersion $(k_d$ vs. $k'_d)$ amongst the homozygous loci. For example, for any of the k possible homozygotes among second cycle hybrid genotypes $k_d = k$ and $k-k_d = 0$. Therefore $\bar{G}_{ii} = m + [d]_{ii}$ and $\bar{G}'_{ii} = m + [d]_{ii} + [i]_{ii}$. Furthermore, for the most associated high scoring of these homozygotes (P_1) all d's are positive and its mean $\bar{P}_1 = \bar{G} = m + \sum^k d$ or $\bar{P}_1 = \bar{G}' = m + \sum^k d +$

$\sum^{(1/2)k(k-1)} i$ according to the prevailing gene action or interaction. Similarly for the most associated low scoring homozygote (P_s) the two expectations are $\bar{P}_s = \bar{G} = m - \sum^k d$ and $\bar{P}_s = \bar{G}' = m - \sum^k d + \sum^{(1/2)k(k-1)} i$ respectively.

On the other hand, the expectation of the original F_1 is obtained when $k_d = 0$ and $k - k_d = k$. Here $\bar{F}_1 = \bar{G} = m + \sum^k h$ or $\bar{F}_1 = \bar{G}' = m + \sum^k h + \sum^{(1/2)k(k-1)} l$ for the non-interacting and interaction models, respectively. The novel heterozygotes among the second cycle hybrids, however, are defined by all other combinations of k_d, k'_d and $(k-k_d)$ within the simultaneous conditions $1 \leqq k_d < k$ and $1 \leqq (k-k_d) < k$.

THEORETICAL CONSIDERATIONS ON THE DATA

The second cycle hybrids can be obtained by crossing the first cycle recombinant inbreds in a diallel fashion, as described in previous work by the authors. The same set of genotypes are also produced by a much less laborious procedure of crossing each of these inbred lines to the F_1 of the original cross. Hence their expectations are the same irrespective of the nature of the gene action and interaction. Table 1 shows the genetical expectations and frequencies of the individual genotypes and the mean and variance of each half sib family (array) obtained by crossing each first cycle inbred line to the F_1 of the initial cross, for two non-interacting loci. They are exactly the same as the value and frequency of individual genotypes and array means and variances obtained in a conventional diallel set of crosses between inbred lines where $u_a = v_a = u_b = v_b = \frac{1}{2}$ (see Mather and Jinks, 1982, for a recent account). This identity holds for any number of loci. The main difference, however, between the two ways of obtaining the random sample of second cycle hybrids is that the diallel consists of genetically homogeneous families which can be replicated whereas the equivalent produced by crossing the inbred lines to the F_1 consists of unique unreplicable individuals. Therefore, as long as we confine our interest to the proportions of genotypes that exceed or fall short of some chosen standards and do not make assertions about specific hybrid genotypes, the data from the inbred lines × F_1 is an adequate substitute for the diallel for testing the validity of our predictions. Because, however, we cannot replicate the hybrid genotypes the phenotypic distributions will have an environmental component which will be substantial for characters of low heritability. In making our predictions for comparison with these

Table 1 Genetical expectations of the individuals, family means and within family variances of first cycle inbred $\times F_1$ crosses on a simple additive and dominance model

Genotype of the inbred	Frequency	F_1 Gametes AB $\frac{1}{4}$	Ab $\frac{1}{4}$	aB $\frac{1}{4}$	ab $\frac{1}{4}$	Family mean	Within family variance
$AABB$	$\frac{1}{4}$	*$AABB$ †$d_a + d_b$	$AABb$ $d_a + h_b$	$AaBB$ $h_a + d_b$	$AaBb$ $h_a + h_b$	$m + \frac{1}{2}(d_a + d_b) + \frac{1}{2}(h_a + h_b)$	$\frac{1}{4}(d_a^2 + d_b^2) + \frac{1}{4}(h_a^2 + h_b^2)$ $- \frac{1}{2}(d_a h_a + d_b h_b)$
$AAbb$	$\frac{1}{4}$	$AABb$ $d_a + h_b$	$AAbb$ $d_a - d_b$	$AaBb$ $h_a + h_b$	$Aabb$ $h_a - d_b$	$m + \frac{1}{2}(d_a - d_b) + \frac{1}{2}(h_a + h_b)$	$\frac{1}{4}(d_a^2 + d_b^2) + \frac{1}{4}(h_a^2 + h_b^2)$ $+ \frac{1}{2}(d_a h_a - d_b h_b)$
$aaBB$	$\frac{1}{4}$	$AaBB$ $h_a + d_b$	$AaBb$ $h_a + h_b$	$aaBB$ $-d_a + d_b$	$aaBb$ $-d_a + h_b$	$m + \frac{1}{2}(-d_a + d_b) + \frac{1}{2}(h_a + h_b)$	$\frac{1}{4}(d_a^2 + d_b^2) + \frac{1}{4}(h_a^2 + h_b^2)$ $+ \frac{1}{2}(-d_a h_a + d_b h_b)$
$aabb$	$\frac{1}{4}$	$AaBb$ $h_a + h_b$	$Aabb$ $h_a - d_b$	$aaBb$ $-d_a + h_b$	$aabb$ $-d_a - d_b$	$m + \frac{1}{2}(-d_a - d_b) + \frac{1}{2}(h_a + h_b)$	$\frac{1}{4}(d_a^2 + d_b^2) + \frac{1}{4}(h_a^2 + h_b^2)$ $+ \frac{1}{2}(d_a h_a + d_b h_b)$

* genotypes
† genotypic deviations from m

phenotypic distributions, we must, therefore, use the formulae that allow for this environmental component (Toledo *et al.*, 1984). At the same time, since we are also interested in the expected properties of the true F_1 hybrids that would be produced in a highly replicated diallel set of crosses we shall also make the predictions by the usual formulae which omit the environmental component.

Another possible source of differences between the random sample of hybrids from the inbred lines $\times F_1$ and the diallel set of crosses is linkage. This is because a linkage disequilibrium will be

Table 2 Genotypic values, their frequencies and family (array) means for the inbred line $\times F_1$ crosses on a linked (coupling phase) digenic interaction mdel

Genotype of the inbred	Frequency	F_1 Gametes AB $\frac{1-p}{2}$	Ab $\frac{p}{2}$	aB $\frac{p}{2}$	ab $\frac{1-p}{2}$	Family (array) mean†
$AABB$	$\frac{1}{2}\left(\frac{1}{1+2p}\right)$	$AABB*$ $\left(\frac{1-p}{1+2p}\right)$ $d_a + d_b + i_{ab}$	$AABb$ $\left(\frac{p}{1+2p}\right)$ $d_a + h_b + j_{ab}$	$AaBB$ $\left(\frac{p}{1+2p}\right)$ $h_a + d_b + j_{ba}$	$AaBb$ $\left(\frac{1-p}{1+2p}\right)$ $h_a + h_b + l_{ab}$	$\frac{1}{1+2p}(d_a + d_b) + \frac{1}{1+2p}(h_a + h_b) + \left(\frac{1-p}{1+2p}\right)\iota$ $+ \frac{p}{1+2p}(j_{ab} + j_{ba}) + \frac{1-p}{1+2p} l_{ab}$
$AAbb$	$\frac{1}{2}\left(\frac{2p}{1+2p}\right)$	$AABb$ $\left(\frac{2p(1-p)}{1+2p}\right)$ $d_a + h_b + j_{ab}$	$AAbb$ $\left(\frac{2p^2}{1+2p}\right)$ $d_a - d_b - i_{ab}$	$AaBb$ $\left(\frac{2p^2}{1+2p}\right)$ $h_a + h_b + l_{ab}$	$Aabb$ $\left(\frac{2p(1-p)}{1+2p}\right)$ $h_a - d_b - j_{ba}$	$\frac{2p}{1+2p}(d_a - d_b) + \frac{2p}{1+2p}(h_a + h_b) - \frac{2p^2}{1+2p} i_{ab}$ $+ \frac{2p(1-p)}{1+2p}(j_{ab} - j_{ba}) + \frac{2p^2}{1+2p} l_{ab}$
$aaBB$	$\frac{1}{2}\left(\frac{2p}{1+2p}\right)$	$AaBB$ $\left(\frac{2p(1-p)}{1+2p}\right)$ $h_a + d_b + j_{ba}$	$AaBb$ $\left(\frac{2p^2}{1+2p}\right)$ $h_a + h_b + l_{ab}$	$aaBB$ $\left(\frac{2p^2}{1+2p}\right)$ $-d_a + d_b - i_{ab}$	$aaBb$ $\left(\frac{2p(1-p)}{1+2p}\right)$ $-d_a + h_b - j_{ba}$	$\frac{2p}{1+2p}(-d_a + d_b) + \frac{2p}{1+2p}(h_a + h_b) - \frac{2p^2}{1+2p} i_{ab}$ $- \frac{2p(1-p)}{1+2p}(j_{ab} - j_{ba}) + \frac{2p}{1+2p} l_{ab}$
$aabb$	$\frac{1}{2}\left(\frac{1}{1+2p}\right)$	$AaBb$ $\left(\frac{(1-p)}{1+2p}\right)$ $h_a + h_b + l_{ab}$	$Aabb$ $\left(\frac{p}{1+2p}\right)$ $h_a - d_b - j_{ba}$	$aaBb$ $\left(\frac{p}{1+2p}\right)$ $-d_a + h_b - j_{ab}$	$aabb$ $\left(\frac{1-p}{1+2p}\right)$ $-d_a - d_b + i_{ab}$	$\frac{1}{1+2p}(-d_a - d_b) + \frac{1}{1+2p}(h_a + h_b) + \frac{1-p}{1+2p} i_{ab}$ $- \frac{p}{1+2p}(j_{ab} + j_{ba}) + \frac{1-p}{1+2p} l_{ab}$

* frequencies should be divided by 4
† family means should be divided by 4

greater in the original F_1 than in the inbred lines that are extracted from it by single seed descent or pedigree inbreeding. The effect of linkage will therefore be greater in the inbred line $\times F_1$ crosses than in the diallel set of crosses. Tables 2 and 3 show the genotypic values, their frequencies and array means for the inbred lines $\times F_1$ and the diallel set of crosses respectively, for two loci showing non-allelic interactions and coupling linkage. For repulsion linkage the frequencies of lines and of gametes in the coupling and repulsion configurations must be interchanged. This has the consequence of changing the sign of the linkage bias from plus to minus. However, in practice it is likely that some loci will be unlinked, some will be linked in coupling and others in repulsion. The net contributions of the two linkage phases will therefore determine the overall linkage bias and hence the differences between the two ways of generating random samples of the second cycle hybrids.

MATERIALS AND METHODS

The experimental data comes from generations derived from the $V_1 \times V_5$ and $V_2 \times V_{12}$ crosses of *Nicotiana rustica*. The experiment involving the material from the $V_1 \times V_5$ cross was grown in 1973 and included the six basic generations, a set of 82 first cycle recombinant inbred lines and the L_1, L_2 and L_3 generations of an inbred line triple test cross (Pooni and Jinks, 1976; Pooni, Jinks and Jayasekara, 1978). From the six basic generations, estimates of the genetical parameters m, $[d]$, $[h]$, $[i]$, $[j]$ and $[l]$ were obtained along with estimates of D as $[4V_{F_2} - 2(V_{B_1} + V_{B_2})]$ and E as $[\frac{1}{2}V_{F_1} + \frac{1}{4}(V_{P_1} + V_{P_2})]$ (Mather and Jinks, 1982). Predictions concerning the proportions of second cycle hybrids scoring $\geq \bar{P}_1$, $\leq \bar{P}_2$, $\geq \bar{F}_1$ and $\geq \bar{P}_l$ and $\leq \bar{P}_s$ the most associated first cycle, recombinant inbred lines were made for the characters height at two, four and six weeks ($H2$, $H4$ and $H6$, respectively), flowering time (FT), height at flowering time (HFT), leaf length and width (LL and LW, respectively) and final height (FH). Similarly, predictions concerning the proportion of first cycle inbred lines scoring higher or lower than the standards referred to earlier were also made. Two sets of predictions for the hybrids were made. The first, which allows for the environmental component and predicts the phenotypic distribu-

Table 3 Genetical expectations of a diallel set of crosses produced by crossing a random sample of first cycle inbreds for a pair of loci showing linked (coupling phase) digenic interactions

Genotype of the inbreds	Frequency	AABB $\frac{1}{2}\left(\frac{1}{1+2p}\right)$	AAbb $\frac{1}{2}\left(\frac{2p}{1+2p}\right)$	aaBB $\frac{1}{2}\left(\frac{2p}{1+2p}\right)$	aabb $\frac{1}{2}\left(\frac{1}{1+2p}\right)$	Array mean†
		AABB	AABb	AaBB	AaBb	
AABB	$\frac{1}{2}\left(\frac{1}{1+2p}\right)$	$*\left(\frac{1}{(1+2p)^2}\right)$ $d_a + d_b + i_{ab}$	$\left(\frac{2p}{(1+2p)^2}\right)$ $d_a + h_b + j_{ab}$	$\left(\frac{2p}{(1+2p)^2}\right)$ $h_a + d_b + j_{ba}$	$\left(\frac{1}{(1+2p)^2}\right)$ $h_a + h_b + l_{ab}$	$\frac{1}{1+2p}(d_a + d_b) + \frac{1}{1+2p}(h_a + h_b) + \left(\frac{1}{1+2p}\right)^2 i_{ab} + \frac{2p}{(1+2p)^2}(j_{ab} + j_{ba}) + \frac{1}{(1+2p)^2}l_{ab}$
		AABb	AAbb	AaBb	Aabb	
AAbb	$\frac{1}{2}\left(\frac{2p}{1+2p}\right)$	$\left(\frac{2p}{(1+2p)^2}\right)$ $d_a + h_b + j_{ab}$	$\left(\frac{2p}{1+2p}\right)^2$ $d_a - d_b - i_{ab}$	$\left(\frac{2p}{1+2p}\right)^2$ $h_a + h_b + l_{ab}$	$\left(\frac{2p}{(1+2p)^2}\right)$ $h_a - d_b - j_{ba}$	$\frac{2p}{1+2p}(d_a - d_b) + \frac{2p}{1+2p}(h_a + h_b) - \left(\frac{2p}{1+2p}\right)^2 i_{ab} + \frac{2p}{(1+2p)^2}(j_{ab} - j_{ba}) + \left(\frac{2p}{1+2p}\right)^2 l_{ab}$
		AaBB	AaBb	aaBB	aaBb	
aaBB	$\frac{1}{2}\left(\frac{2p}{1+2p}\right)$	$\left(\frac{2p}{(1+2p)^2}\right)$ $h_a + d_b + j_{ba}$	$\left(\frac{2p}{1+2p}\right)^2$ $h_a + h_b + l_{ab}$	$\left(\frac{2p}{1+2p}\right)^2$ $-d_a + d_b - i_{ab}$	$\left(\frac{2p}{(1+2p)^2}\right)$ $-d_a + h_b - j_{ba}$	$\frac{2p}{1+2p}(-d_a + d_b) + \frac{2p}{1+2p}(h_a + h_b) - \left(\frac{2p}{1+2p}\right)^2 i_{ab} - \frac{2p}{(1+2p)^2}(j_{ab} - j_{ba}) + \left(\frac{2p}{1+2p}\right)^2 l_{ab}$
		AaBb	Aabb	aaBb	aabb	
aabb	$\frac{1}{2}\left(\frac{1}{1+2p}\right)$	$\left(\frac{1}{(1+2p)^2}\right)$ $h_a + h_b + l_{ab}$	$\left(\frac{1}{(1+2p)^2}\right)$ $h_a - d_b - j_{ba}$	$\left(\frac{2p}{(1+2p)^2}\right)$ $-d_a + h_b - j_{ab}$	$\left(\frac{1}{(1+2p)^2}\right)$ $-d_a - d_b + i_{ab}$	$\frac{1}{1+2p}(-d_a - d_b) + \frac{1}{1+2p}(h_a + h_b) + \left(\frac{1}{1+2p}\right)^2 i_{ab} - \frac{2p}{(1+2p)^2}(j_{ab} + j_{ba}) + \frac{1}{(1+2p)^2}l_{ab}$

* frequencies should be divided by 4
† array means should be divided by 4

tion for comparison with the observed distribution of the L_3 generation

$$c = \frac{St - \bar{F}_2}{(V_{1F_2})^{1/2}},$$

and the second, which omits the environmental component and predicts the genotypic distribution of the true means of the second cycle F_1 hybrids

$$c = \frac{St - \bar{F}_2}{(_H V_{1F_2})^{1/2}},$$

where St is the mean of any of the chosen standards and $\bar{F}_2$, V_{1F_2} and $_H V_{1F_2}$ are the mean and the phenotypic and heritable variance of the F_2 generation, respectively. The predictions for the inbred lines were made using the formula

$$c = \frac{St - m}{(D)^{1/2}},$$

where St, m and D are as defined earlier (Mather and Jinks, 1982; Jinks and Perkins, 1972; Jinks and Pooni, 1976; Pooni and Jinks, 1978, 1979). The observed proportions of second cycle hybrids for various categories ($\geq \bar{P}_1$, $\leq \bar{P}_2$, $\geq \bar{F}_1$, $\geq \bar{P}_t$ and $\leq \bar{P}_s$) were obtained by comparing the scores of individuals of the inbred lines $\times F_1$ crosses (L_3 generation) with the means of each of the standard genotypes for each character. The predictions and observations, however, were made in a single experiment, a situation not likely to occur in practice.

The experiments involving material from the $V_2 \times V_{12}$ cross were raised in 1974 and 1975. In the 1974 experiment, the six basic generations of the cross were grown along with the first cycle recombinant inbreds D_{10}, D_{17} and the F_1, F_2, B_1 and B_2 generations derived from the cross $D_{10} \times D_{17}$. Lines D_{10} and D_{17} are the most associated pair of inbreds amongst the 60 D inbred lines extracted from the $V_2 \times V_{12}$ cross for final height (FH). They are also amongst the most highly associated inbreds for the characters FT and HFT (Pooni and Jinks, 1982). The 1975 experiment involved V_2, V_{12}, D_{10}, D_{17}, 60 inbred lines and two sets of L_1, L_2 and L_3 generations of the inbred line $TTCs$ where the testers used were V_2, V_{12} and $V_2 \times V_{12}$ and D_{10}, D_{17} and $D_{10} \times D_{17}$, respectively. Predictions were made from the 1974 experiment and tested by observations from the 1975 experiment following the same procedures described earlier for the $V_1 \times V_5$ cross. For the $V_2 \times V_{12}$ cross, however, data were not available for the characters LL and LW and there was no suitable associated P_t and P_s genotypes for the characters H_2, H_4 and H_6. On the other hand, the availability of the two L_3 generations for this material involving the original and second cycle F_1 allows us to assess the extent to which linkage disequilibrium may bias our results. The two L_3 distributions are expected to differ and these differences are expected to be more than those between the L_3 involving the original F_1 and the diallel set of crosses. Differences between the two L_3 distributions, therefore, provide an indicator of possible differences between the diallel and the L_3 involving the original F_1 which we are substituting for it.

RESULTS AND CONCLUSIONS

The means, variances and numbers of individuals scored for each character for the P_1, P_2, F_1, F_2, B_1, B_2, P_t and P_s families of the $V_1 \times V_5$ and $V_2 \times V_{12}$ crosses are given in table 4. Weighted least squares estimates of the components of means m, $[d]$, $[h]$, $[i]$, $[j]$ and $[l]$ and the components of variances D, E and $_H V_{1F_2}$ are given in table 5. These estimates of statistics and components convey the wide diversity of gene action and interaction and environmental involvement provided by the characters we have studied in the two crosses.

The estimates in tables 4 and 5 plus $V_{1F_2}(= V_{H1F_2} + E)$ have been used to predict the transgressiveness of the phenotypic and genotypic distributions of the second cycle hybrids using the formulae in "Materials and Methods". These predictions are given in table 6. The present experiment provides corresponding observations on the phenotypic distributions only and these are obtained by counting all of the phenotypes in the L_3 families whose scores fall within the categories specified in table 6 and expressing them as a proportion of the total number of individuals. These proportions are given in table 6 for comparison with the predictions.

We have no observations on the genotypic distributions for comparison with the genotypic predictions in table 6. These can be estimated from the observed phenotypic distributions by the method of Robson, Powers and Urquhart (1967) but we prefer to await the outcome of current experiments on a replicated full diallel set of second cycle hybrids.

The standard genotypic predictions for the first cycle recombinant inbred lines for the two crosses are given in table 7 for comparison.

The predictions and observations in tables 6 and 7 provide a wide spectrum of the possible situations a breeder may encounter in practice.

Table 4 Family means and within Generation variances of the six basic generations and the highest and lowest scoring first cycle inbreds (P_l and P_s) derived from the $V_1 \times V_5$ (1973 experiment) and $V_2 \times V_{12}$ (1974 experiment) crosses of *Nicotiana rustica*

Character	P_1	P_2	F_1	F_2	B_1	B_2	P_l	P_s
A. Generation Means								
(*i*) $V_1 \times V_5$ *cross*								
H2	6·8	4·6	6·7	5·6	6·6	5·1	9·2	2·7
H4	19·5	13·1	19·1	15·8	19·1	14·5	26·4	6·3
H6	55·0	46·6	59·2	50·2	56·0	48·0	64·5	22·3
FT	15·3	10·2	11·7	14·1	15·3	11·6	31·3	5·3
HFT	75·9	60·8	72·8	73·7	78·1	68·5	110·6	49·4
LL	17·5	16·9	19·3	18·9	18·9	18·4	23·1	12·4
LW	17·0	15·0	18·1	17·5	18·0	16·6	20·1	11·3
FH	127·3	103·1	130·3	126·5	132·3	118·2	143·8	92·3
n	120	140	140	240	160	200	10	10
(*ii*) $V_2 \times V_{12}$ *cross*								
H2	6·3	2·4	4·8	4·3	5·5	3·5	—	—
H4	25·1	6·2	17·7	15·1	21·4	11·1	—	—
H6	73·1	21·0	60·2	49·0	66·7	36·4	—	—
FT	56·9	38·3	42·2	45·5	49·5	40·8	59·2	35·7
HFT	79·4	68·1	86·3	83·0	85·2	79·1	124·9	58·6
FH	125·4	103·5	158·7	140·7	143·6	136·3	167·4	76·0
n	40	20	236	626	336	698	40	20
B. Generation Variances								
(*i*) $V_1 \times V_5$ *cross*								
H2	6·2	3·7	6·1	6·4	6·9	5·4	2·9	1·0
H4	53·4	30·5	48·6	53·0	53·6	39·1	22·9	6·1
H6	174·0	194·3	237·8	268·0	220·9	246·8	54·7	85·9
FT	29·8	22·0	23·0	40·0	34·7	26·8	121·5	4·2
HFT	250·4	50·0	159·8	218·6	269·2	119·7	298·7	77·4
LL	6·4	5·7	6·2	8·7	7·1	7·0	1·5	3·6
LW	7·9	7·6	8·9	10·2	8·7	9·0	6·3	5·5
FH	303·0	95·6	229·7	265·6	264·6	222·8	346·3	135·0
(*ii*) $V_2 \times V_{12}$ *cross*								
H2	1·9	0·3	1·2	3·0	2·3	1·4	—	—
H4	24·2	1·4	19·5	33·0	25·8	14·5	—	—
H6	86·3	21·9	88·2	209·9	117·9	114·0	—	—
FT	41·7	7·1	13·8	47·2	53·2	14·4	32·7	5·3
HFT	63·3	165·0	148·4	361·9	145·3	325·7	182·7	9·9
FH	121·4	92·7	186·1	479·1	196·9	509·9	384·8	54·1

Table 5 Estimates of components of generation means and variances for the $V_1 \times V_5$ and $V_2 \times V_{12}$ cross of *Nicotiana rustica*

Character	*m*	[*d*]	[*h*]	[*i*]	[*j*]	[*l*]	$\chi^2(df)$	D	$_H V_{1F2}$	E
				Components of means					Components of Variances	
				(*i*) $V_1 \times V_5$ *cross*						
H2	4·45 ± 0·38	1·19 ± 0·12	2·21 ± 0·53	1·27 ± 0·41	—	—	2·11 (2)*	1·06	0·91	5·53
H4	12·42 ± 1·07	3·53 ± 0·35	6·63 ± 1·50	3·97 ± 1·16	—	—	2·86 (2)	22·64	7·73	45·28
H6	40·76 ± 2·39	5·02 ± 0·74	18·17 ± 3·34	9·95 ± 2·57	—	—	4·81 (2)	136·28	57·01	210·95
FT	16·74 ± 0·85	2·81 ± 0·28	−4·96 ± 1·13	−3·92 ± 0·92	—	—	3·55 (2)	36·84	15·50	24·47
HFT	73·75 ± 0·57	7·99 ± 0·67	—	−5·02 ± 0·95	—	—	3·11 (3)	96·60	61·36	157·22
LL	19·10 ± 0·11	0·35 ± 0·13	—	−1·89 ± 0·20	—	—	2·47 (3)	6·48	2·52	6·13
LW	16·12 ± 0·16	1·09 ± 0·15	2·25 ± 0·29	—	—	—	4·69 (3)	5·48	1·87	8·33
FH	123·08 ± 2·37	12·57 ± 0·76	7·44 ± 3·29	−7·58 ± 2·55	—	—	1·49 (2)	87·56	51·09	214·50
				(*ii*) $V_2 \times V_{12}$ *cross*						
H2	3·85 ± 0·14	1·97 ± 0·74	0·97 ± 0·19	0·55 ± 0·19	—	—	0·84 (2)	4·62	1·83	1·16
H4	12·64 ± 0·52	9·96 ± 0·25	5·15 ± 0·73	3·48 ± 0·62	—	—	3·46 (2)	51·44	16·84	16·16
H6	38·00 ± 1·23	26·11 ± 0·88	22·28 ± 1·66	9·21 ± 1·58	8·23 ± 2·20	—	0·20 (1)	375·64	138·70	71·15
FT	47·91 ± 0·31	9·05 ± 0·29	−5·52 ± 0·46	—	—	—	5·19 (3)	53·46	28·10	19·10
HFT	80·28 ± 1·61	5·84 ± 0·79	6·24 ± 2·16	−6·21 ± 2·13	—	—	1·02 (2)	505·56	230·61	131·28
FH	114·88 ± 1·34	8·79 ± 0·88	57·13 ± 4·53	—	—	−13·31 ± 3·74	4·84 (2)	502·66	332·49	146·60

* each χ^2 value has $P > 5$ per cent.

Table 6 Predicted and observed proportions of second cycle hybrids for $V_1 \times V_5$ and $V_2 \times V_{12}$ crosses of *Nicotiana rustica*

Characters		$\geq \bar{P}_1$ Pred	Obs	$\leq \bar{P}_2$ Pred	Obs	Categories $\geq \bar{F}_1$ Pred	Obs	$\geq \bar{P}_l$ Pred	Obs	$\leq \bar{P}_s$ Pred	Obs
(i) $V_1 \times V_5$ *cross*											
H2	*G	0·10		0·15		0·13		0·00		0·00	
	†Ph	0·32	0·21	0·35	0·54	0·33	0·21	0·08	0·04	0·13	0·14
H4	G	0·09		0·17		0·12		0·00		0·00	
	Ph	0·31	0·20	0·36	0·53	0·33	0·20	0·07	0·04	0·10	0·10
H6	G	0·26		0·32		0·12		0·03		0·00	
	Ph	0·39	0·31	0·41	0·51	0·29	0·22	0·19	0·12	0·04	0·08
FT	G	0·38		0·16		0·73		0·00		0·01	
	Ph	0·42	0·42	0·27	0·19	0·65	0·75	0·00	0·02	0·08	0·03
HFT	G	0·39		0·05		0·54		0·00		0·00	
	Ph	0·44	0·40	0·19	0·20	0·52	0·48	0·01	0·02	0·05	0·02
LL	G	0·81		0·10		0·40		0·00		0·00	
	Ph	0·68	0·77	0·25	0·18	0·44	0·52	0·08	0·07	0·01	0·01
LW	G	0·64		0·03		0·33		0·03		0·00	
	Ph	0·56	0·63	0·22	0·18	0·42	0·46	0·21	0·23	0·03	0·02
FH	G	0·46		0·00		0·30		0·01		0·00	
	Ph	0·48	0·55	0·07	0·08	0·41	0·48	0·14	0·21	0·02	0·02
(ii) $V_2 \times V_{12}$ *cross*											
H2	G	0·07		0·08		0·36		—		—	
	Ph	0·12	0·10	0·14	0·26	0·39	0·37				
H4	G	0·01		0·01		0·26		—		—	
	Ph	0·04	0·06	0·06	0·04	0·33	0·31				
H6	G	0·02		0·01		0·17		—		—	
	Ph	0·05	0·03	0·03	0·03	0·22	0·24				
FT	G	0·02		0·09		0·73		0·01		0·03	
	Ph	0·05	0·12	0·15	0·19	0·68	0·65	0·02	0·04	0·08	0·10
HFT	G	0·60		0·16		0·41		0·00		0·05	
	Ph	0·58	0·38	0·22	0·18	0·43	0·43	0·01	0·01	0·10	0·28
FH	G	0·80		C·02		0·16		0·07		0·00	
	Ph	0·76	0·63	0·04	0·04	0·21	0·28	0·11	0·05	0·00	0·01

* *G* is genotypic predictions
† *Ph* is phenotypic predictions and observations

Table 7 Predicted proportions (genotypic) of first cycle inbred lines which can be extracted by random inbreeding of the $V_1 \times V_5$ and $V_2 \times V_{12}$ crosses of *Nicotiana rustica*

Characters	$\geq \bar{P}_1$	$\leq \bar{P}_2$	$\geq \bar{F}_1$	$\geq \bar{P}_l$	$\leq \bar{P}_s$
(i) $V_1 \times V_5$ *cross*					
H2	0·01	0·56	0·01	0·00	0·04
H4	0·07	0·56	0·08	0·00	0·10
H6	0·11	0·69	0·06	0·02	0·06
FT	0·59	0·14	0·80	0·08	0·03
HFT	0·41	0·09	0·54	0·00	0·01
LL	0·74	0·19	0·47	0·06	0·00
LW	0·35	0·32	0·20	0·04	0·02
FH	0·33	0·02	0·22	0·01	0·00
(ii) $V_2 \times V_{12}$ *cross*					
H2	0·13	0·25	0·33	—	—
H4	0·04	0·18	0·24	—	—
H6	0·04	0·19	0·13	—	—
FT	0·11	0·10	0·78	0·06	0·05
HFT	0·52	0·29	0·39	0·02	0·17
FH	0·32	0·31	0·03	0·01	0·04

They confirm (see table 7) the now well established ease with which first cycle recombinant inbreds can be extracted from a cross which outperform their parents and even the most heterotic F_1's (Jinks and Pooni, 1976; Jinks, 1981, 1983). They also show that there is a small probability of improving upon the extremes of the first cycle recombinant inbred lines (P_l and P_s) by either second cycle F_1's (table 6) or second cycle recombinant inbreds derived from them (table 7).

The predicted and observed proportions of second cycle hybrids falling into the various categories (table 6) show with few exceptions very good agreement. As expected, the phenotypic predictions are closer to the observed proportions than the genotypic predictions, which, as usual, are generally conservative thus confirming that the phenotypic predictions are more appropriate when the observed distributions have a substantial environmental component. The overall correlations between the observed and predicted

properties are 0·81 and 0·88 for the $V_1 \times V_5$ and $V_2 \times V_{12}$ crosses, respectively. While this good agreement between the phenotypic predictions and the observed proportions validates our procedures it must not be forgotten that the potential to produce second cycle hybrids which consistently outperform the standards is given by the more conservative genotypic predictions in table 6.

The origin of the second cycle F_1 hybrids which we have predicted and observed to score higher (or lower) than the corresponding extremes, P_l (or P_s), among the first cycle recombinant inbred lines or the original F_1, can be clarified by the general expectations developed in section 2. For an additive-dominance model of gene action the generalised second cycle F_1 mean is

$$\bar{G} = m + \sum^{k_d} d - 2 \sum^{k'_d} d + \sum^{k-k_d} h.$$

If P_l were the completely associated inbred line, that is $P_l = m + \sum^k d$, no second cycle F_1 could have a higher score without overdominance $(\sum^k h > \sum^k d)$. If, for reasons we shall discuss later, P_l is not completely associated because of decreasing alleles at k'_d loci then

$$\bar{P}_l = m + \sum^{k_d} d - 2 \sum^{k'_d} d (= \bar{G} \text{ when } k_d = k)$$

and a second cycle F_1 will have a higher score whenever there is heterozygosity at the k'_d loci, irrespective of the direction of the dominance deviations (h positive or negative) providing that, if the decreasing allele is dominant (h negative), dominance is incomplete. On the other hand, second cycle F_1's will have higher scores than the original F_1 whose expected mean is $m + \sum^k h$ whenever homozygosity for the increasing allele is substituted at loci at which there is either incomplete dominance ($h < d$) or dominance for the decreasing allele (h negative). It would also occur if homozygosity for the decreasing allele were substituted at loci at which there is overdominance for the decreasing allele (h negative $|h| > d$).

Since dominance is incomplete for all characters in the $V_1 \times V_5$ and $V_2 \times V_{12}$ crosses (Pooni, Jinks and Jayasekara, 1978) it is clear that the highest scoring second cycle F_1's are most likely to arise in crosses among first cycle recombinant inbred lines which have a high proportion of increasing alleles, that is $k_d - k'_d$ approaches k. As a consequence these second cycle F_1's have very similar genotypes to the highest scoring second cycle inbreds (Toledo, Pooni and Jinks, 1984). The second cycle F_1's of the L_3 generation confirm this expectation in that the higher the score

of the first cycle recombinant inbred line the higher the mean of the second cycle F_1 progeny it produces.

Since, in the absence of linkage, the frequency of the highest scoring completely associated inbred line that can be extracted from a cross is $(\frac{1}{2})^k$, unless k is small (<9) we cannot expect to find such a line even in the largest collections so far examined (Ingram and Jinks, 1982, 472 inbred lines; Jinks and Pooni, 1984, 844 inbred lines). In the present study with sample sizes of 60 and 82 inbred lines the probability of finding the highest scoring completely associated line will be very low if k is greater than 6. Where, as in this study, there is a linkage disequilibrium due to an excess of repulsion linkages (Jinks and Perkins, 1969; Perkins and Jinks, 1970; Jinks and Pooni, 1981, 1982, 1984) the probability of recovering the completely associated inbred line will be even lower. Hence the probability of obtaining second cycle F_1's with a higher mean than the highest inbred line in samples of 60 and 82 will be correspondingly increased.

If we further introduce the complication of non-allelic interactions whose presence is well established in the material used in this study (Pooni, Jinks and Jayasekara, 1978) there are many more options leading to second cycle F_1's which outperform the best of the first cycle recombinant inbred lines as we shall show in a later paper.

REFERENCES

INGRAM, N. R. AND JINKS, J. L. 1982. Analysis of induced quantitative variation. II. The recombinant inbred lines. *Heredity*, 48, 79–83.

JINKS, J. L. 1981. The genetical framework of plant breeding. *Phil. Trans. Roy. Soc. Lond.*, B 292, 407–419.

JINKS, J. L. 1983. Biometrical Genetics of Heterosis, In *Heterosis.* Monographs on theoretical and applied genetics 6. R. Frankel (ed.). Springer-Verlag, Berlin, 1–46.

JINKS, J. L. AND PERKINS, J. M. 1969. The detection of linked epistatic genes for a metrical trait. *Heredity*, 24, 465–475.

JINKS, J. L. AND PERKINS, J. M. 1972. Predicting the range of inbred lines. *Heredity*, 28, 399–403.

JINKS, J. L. AND POONI, H. S. 1976. Predicting the properties of recombinant inbred lines derived by single seed descent. *Heredity*, 36, 253–266.

JINKS, J. L. AND POONI, H. S. 1981. Properties of pure-breeding lines produced by dihaploidy, single seed descent and pedigree breeding. *Heredity*, 46, 391–395.

JINKS, J. L. AND POONI, H. S. 1982. Predicting the properties of pure breeding lines extractable from a cross in the presence of linkage. *Heredity*, 49, 265–270.

JINKS, J. L. AND POONI, H. S. 1984. Comparison of inbred lines produced by single seed descent and pedigree inbreeding. *Heredity*, in press.

MATHER, K. AND JINKS, J. L. 1982. *Biometrical Genetics*, 3rd Edn. Chapman and Hall, London.

PERKINS, J. M. AND JINKS, J. L. 1970. Detection and estimation of genotype-environmental, linkage and epistatic components of variation for a metrical trait. *Heredity*, *25*, 157–177.

POONI, H. S. AND JINKS, J. L. 1976. The efficiency and optimal size of triple test cross designs for detecting epistatic variation. *Heredity*, *36*, 215–227.

POONI, H. S. AND JINKS, J. L. 1978. Predicting the properties of recombinant inbred lines derived by single seed descent for two or more characters simultaneously. *Heredity*, *40*, 349–361.

POONI, H. S. AND JINKS, J. L. 1979. Sources and biases of the predictors of the properties of recombinant inbreds produced by single seed descent. *Heredity*, *42*, 41–48.

POONI, H. S. AND JINKS, J. L. 1982. Comparative analysis of association and dispersion crosses to detect linkage and epistatic components of variation. *Heredity*, *49*, 211–220.

POONI, H. S., JINKS, J. L. AND JAYESEKARA, N. E. M. 1978. An investigation of gene action and genotype × environment interactions in two crosses of *Nicotiana rustica* by triple test cross and inbred line analysis. *Heredity*, *41*, 83–92.

ROBSON, D. S., POWERS, L. AND URQUHART, N. S. 1967. The proportion of genetic deviates in the tails of a normal population. *Der Züchter*, *37*, 205–216.

TOLEDO, J. F. F. DE, POONI, H. S. AND JINKS, J. L. 1984. Predicting the properties of second cycle hybrids produced by intercrossing random samples of recombinant inbred lines. *Heredity*, *53*, 283–292.

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*, *XXX*, 201–232.