Theor Appl Genet (1989) 77:260–270

Genetic control of seed weight in flax (*Linum usitatissimum*) and possible implications

H. Tyson
Biology Department, McGill University, Montreal, Quebec, Canada

Received May 31, 1988; Accepted September 8, 1988
Communicated by A.R. Hallauer

Summary. Mean seed weight data were obtained from the F_1 and F_2 of a six-by-six diallel cross with flax (*Linum usita-tissimum* L.). Pronounced reciprocal differences appeared in the F_1, but had largely disappeared by the F_2. The genetic control of mean seed weight was examined using two types of analysis of variance. The models underlying both analyses were fitted to the data by matrix methods supplying weighted least-squares estimates of the parameters in the models. Weights, the use of which dealt with the problem of variation in the reliability of means, were the reciprocals of the variances of individual cell (cross/self) means in the diallel data table. Elimination of redundant parameters supplied the minimum adequate models for each analysis type. Dominance was apparently masked by the large transient maternal effects in the F_1, but surfaced in the F_2, where dominance was towards larger mean seed weight. This may be coupled with findings elsewhere of possible advantages for larger seed weights to speculate on a role in preserving infrequent hybrid progeny among inbreeding (flax) species. Maternal effects producing larger seed size, plus dominance with the same result might be valuable, in conjunction with growth and competitive advantages conferred by larger seed, in preventing early elimination of rare hybrids.

Key words: Analyses of seed weight inheritance – Weighted least squares – Flax diallels

Introduction

Few studies have been made on the mode of inheritance of mean seed weight (1,000 K wt.) in flax (*Linum usitatissimum* L.), a completely inbreeding, diploid and commercially exploited species. Smith and Fitzsimmons (1964) found that in a cross of two cultivars differing markedly in seed weight, maternal inheritance appeared in the F_2, together with some indications of simple qualitative inheritance in backcrosses. Jeswani and Murty (1963) suggested that the character is quantitatively inherited, while Khan (1963) showed that 1,000 K weight is inherited additively with little detectable dominance. Where parents differ in characteristic seed weight, transient reciprocal (maternal) effects would be expected in the F_1. A range of seed weight (1,000 K wt.) from 3.2 – 10.6 g is seen among flax genotypes in the World Flax Collection (USDA, Fargo, ND, USA). The correlation between mean seed weight and mean seed volume or size is nearly +1, so that seed weight and size may be used interchangeably.

In the studies on mean seed weight reported here, a diallel cross involving six flax genotypes was analysed. Models to describe the quantitative inheritance shown by this continuously varying character were fitted by least-squares procedures, using weighting to obtain maximum reduction of the residual term after fitting. The models were patterned on those used by Griffing (1956), Hayman (1954a, b), Mather and Jinks (1982) and Morley-Jones (1965) for crosses with two or more than two (diallel) parents. Alternative methods of examining reciprocal differences in detail have been explored by Durrant (1965, 1968), but have not been used here. The objective of the study was to examine the genetic control of seed weight in a diallel cross with F_1 and F_2 generations. The F_1 and F_2 give an opportunity to distinguish transient maternal (reciprocal) effects from more persistent, presumably cytoplasmic factors causing reciprocal differences. With these data, attention is focused on the generation or progeny means; partitioning of generation variances is not dealt with. Conclusions reached with the data reported here refer, strictly speaking, only to these parental geno-

types. Two additional sets of data on mean seed weight from smaller (four parent) diallels were available; these data are examined very briefly here. These two smaller diallels both contain fibre and oilseed flax genotypes and, in one case, two of the parental genotypes in the six-by-six diallel. They also both contain wild genotypes from natural populations, and their data are included to provide a degree of generalization for the main conclusions from the analyses of the larger diallel.

Materials and methods

(a) Parents and crosses

Three diallel crosses ('diallel 1', 'diallel 2', and 'diallel 3') were studied, with some overlap among the parents. All parental genotypes are completely inbreeding; no evidence of outcrossing emerged in the studies reported here or in any previous work. All are diploid, with a range of flower colours, but all produce brown seed, with seed weights varying over the range from 2–9 g per 1,000.

Diallel 1: A six- by six-diallel was completed, which included commercial cultivars (C. I. numbers where available are shown in parentheses) Percello (='P'; C. I. no. 1788), Mandarin (='M'), Royal (='R'; C. I. no. 828) and Dakota (='D'; C. I. no. 1087), plus the large (L) and small (S) flax genotrophs produced by Durrant (1971) from Stormont Cirrus. Plants were sown in spring and grown as spaced plants in rows in field conditions in the U. K. at 52.5°N, 4°W. Three replicates of this diallel were completed, with all parents and reciprocals represented in every replicate. Three replicates were grown for the F_1; six were grown for the F_2 in the following season. Mean seed weight data from the F_1 and F_2 were recorded.

The production and characteristics of the L and S genotrophs have been documented at length (Durrant 1971). Essentially L and S plants display pronounced environmentally induced heritable changes arising within a single original genotype. Initial growth of one generation of this genotype (Stormont Cirrus) in high levels of nitrogen (N), phosphorus (P) and potassium (K), i.e. NPK, induces numerous, pronounced morphological, biochemical and genomic changes compared to one treatment generation in NK. NPK induces morphologically large plants, referred to as the L genotroph, while NK induces small plants, referred to as genotroph S. These changes persist at least 20 generations through the subsequent progenies grown in uniform conditions. The differences have been summarized and reviewed (Fieldes and Tyson 1984).

Cultivars P and M are fibre flax types, typically 1 m high with little side branching and relatively low seed production. R and D are oilseed flax types, selected for high seed production; both are relatively highly branched and shorter than fibre types. The genotype from which L and S originated is a fibre type; S resembles a fibre type in its lack of branching, but L is similar to the oilseed/linseed types in size, branching and seed production.

Diallel 2: A four- by four-diallel with commercial cultivars as parents was studied in a single replicate in field conditions. The parents were the fibre type Stormont Cirrus (='SC'; N. Ireland Flax Breeding Station) and the oilseed types Redwood (='RD'; C. I. no. 1380) and a high seed oil content selection from the Agriculture Canada Research Station, Morden, Manitoba, (='HO'). The fourth parent was an unselected, wild genotype originating from Ethiopia (='A1') and obtained from the USDA World Flax Collection. This parent had been propagated in growth chamber/greenhouse conditions, with complete selfing and the use of single plant progenies over several generations to ensure genetic uniformity and complete homozygosity. A1 is extremely short and very highly branched compared to the other three selections, and its seed weight is also extremely low. F_2 data on seed weight were obtained.

Diallel 3: A four- by four-diallel was studied in growth chambers, using the commercial fibre and oilseed cultivars Royal (R) and Mandarin (M), plus two unselected, wild genotypes originating from Ethiopia (='A2') and from Turkey (='T'). The latter were obtained from the USDA World Flax Collection. A2, like A1, was very short and highly branched, with very small seed. In contrast, T was tall, with very large seed. F_1 data were obtained.

(b) Genetic models for diallel crosses

Analyses of diallel crosses fall into three groups: (1) comparison of array variances and covariances through linear regressions; (2) estimation and testing of general (GCA) and specific (SCA) combining abilities for the parental group; (3) estimation and testing of additive and dominance genetic effects. In the latter two cases, inclusion of reciprocals allows evaluation and testing of reciprocal (maternal) effects. Analyses in (1) may allow confirmation that a simple genetical model which ignores non-allelic interaction is suitable.

The model for estimating SCAs and GCAs has been outlined by Griffing (1956) for observation y_{ijkl} as:

$$y_{ijkl}=\mu+g_i+g_j+s_{ij}+r_{ij}+b_k+bv_{ijk}+e_{ijkl}$$

– where μ is the overall mean, g_i is the GCA effect of the ith parent, s_{ij} is the SCA effect for the cross between parents i and j, r_{ij} is the reciprocal effect for this cross, b_k is the effect of block (replicate) k, bv_{ijk} is the interaction between the ijth genotype and the kth block, and e_{ijkl} is the environmental effect for the $ijkl$th individual. The sums of squares for GCA, SCA and reciprocal effects are computed using scalar formulas detailed by Griffing; the degrees of freedom (*df*) are respectively $(p-1)$, $(p(p-1)/2)$, and $(p(p-1)/2)$, where p is the number of parental genotypes. Griffing does not subdivide the sum of squares for reciprocal effects, which is calculated from the $(p(p-1)/2)$ individual effects, i.e. $r_{ij}=(y_{ij}-y_{ji})/2$. The error term against which the various mean squares are tested is the interaction between blocks (replicates) and the diallel progenies.

Hayman (1954 a, b) developed an analysis of variance which produces estimates and tests the significance of additive (a) and non-additive genetic effects (b), even if reciprocal differences (c, d), which can also be tested, exist among progenies. The model for Hayman's analysis is:

$$y_{rs}=m+j_r+j_s+j_{rs}+k_r-k_s+k_{rs}$$

– where y_{rs} is the progeny of parent r by parents s, m is the overall mean, j_r the mean deviation from the grand due to the rth parent, j_{rs} the remaining discrepancy in the rsth reciprocal sum, $2k_r$ the difference between the effects of the rth parental line used as male parent and as female parent, and $2k_{rs}$ is the remaining discrepancy in the rsth reciprocal difference. The sums of squares accounted for by fitting the parameters j_r, j_{rs}, k_r and k_{rs} represent Hayman's a, b, c and d sums of square above. Griffing's GCA sum of squares is identical to Hayman's additive sum of squares (a's), with identical *df*. The dominance effects (b) in Hayman's analysis are subdivided into: (1) the overall departure of the hybrids from the mean of the parents with 1 *df*, reflecting directional dominance among the loci involved (b_1); (2) the difference among parents, with $(p-1)$ *df*, in the mean dominance

deviation of the hybrids from the mean of that parent's array (b_2); (3) that part of the dominance deviation unique to each hybrid (b_3), with $((p(p-1)/2)-p)\,df$.

For each of these (and the reciprocal effects below), corresponding sums of squares are calculated, leading to significance tests. However, parameter b_3 cannot be made orthogonal to any of the previous parameters, and its sum of squares is simply found by subtraction, given a sum of squares for all three dominance (b) items. The sum of squares for $(b_1+b_2+b_3)$ is equal to Griffing's sum of squares for SCA effects, with identical *df*. The reciprocal differences in Hayman's analysis are subdivided into those for the average maternal effects of each parental line (c), and remaining reciprocal differences (d), with the total sum of squares for *c* and *d* equal to the sums of squares for all individual hybrid reciprocal differences. Hayman's a, b, c and d components are tested against their individual interactions with blocks, interactions which may be pooled if homogeneous to increase the sensitivity of the tests. For both the Griffing and the Hayman-type analyses on the data here, reciprocal effects were simply calculated for individual reciprocal hybrids and summed.

The Griffing and the Hayman models were fitted to the diallel 1 data here by matrix methods, outlined below, which provided weighted least-squares solutions, supplying parameter estimates and their standard errors, plus sums of squares accounted for by individual parameter(s). Weighting progeny means by the reciprocals of the variances of their means for the Griffing, Hayman or other models copes with the frequent heterogeneity among such variances. Variances of means here were computed from individual replicate means for a given cell (three in the F_1, six in the F_2). The need for weighting was established with Bartlett's (1937) test for homogeneity of variances in the F_1 and F_2. There was no evidence here of non-normality in underlying seed weight distributions. The F_1 Bartlett chi square probability was virtually 0.05; the F_2 was highly significant. This heterogeneity implied that F_1 and F_2 weighting would be advantageous. The actual weighting benefits as variances increase in heterogeneity for randomised block experiments are explored separately (Tyson, in preparation). As well as standard significance tests of individual parameters in the model, a weighted solution also allows assessment of the model's adequacy through a chi square test of the residual sum of squares.

(c) Matrix procedures for fitting models

Mather and Jinks (1982) used Cavalli's (1952) technique to fit models to data from a bi-parental cross, where F_1, F_2 and first backcrosses to either parent (B_1, B_2) are available. The simple model for this case, $(y_i=m+[d]+[h])$ is conveniently fitted by matrix methods, requiring (a) description, in terms of a design or $[X]$ matrix, of the particular model to be fitted to the data, and (b) the availability of a suitable computer program dealing with the matrix operations implicit in the weighted least squares solution:

$$\hat{b}=(X'V^{-1}X)^{-1}X'V^{-1}y \qquad (1.1)$$

where $\hat{b}$ is the column vector of parameter estimates, $[X']$ is the transpose of the $[X]$ matrix, $[V^{-1}]$ is a diagonal matrix containing the weights for the generation means, and y is a column vector of the observed generation means. For an unweighted solution, $[V^{-1}]$ is merely the identity matrix. Matrix procedures for a weighted least-squares solution are detailed by Draper and Smith (1968). The sum of squares accounted for by any one or more of the model's parameters in the vector of estimates, b, is (Harvey 1960):

$$\text{Sum of squares}=b'Z^{-1}b \qquad (1.2)$$

where b is now the column vector containing the specific parameter estimates and b' is its transpose, and $[Z]$ is the portion of the $(X'V^{-1}X)^{-1}$ inverse corresponding to the parameter(s) in b for which a sum of squares is required.

A complete three- by three-diallel, with parental genotypes P_1, P_2 and P_3, generates nine progenies:

	P_1	P_2	P_3
P_1	y_{11}	y_{12}	y_{13}
P_2	y_{21}	y_{22}	y_{23}
P_3	y_{31}	y_{32}	y_{33}

— written as nine by one column vector (see below) for use in the matrix solution. The $[X]$ matrix specifies the parameters in the model; for Griffing's (1956) GCA and SCA effects there are 2 *df* for GA effects (g_1, g_2) and 3 *df* for SCA effects (s_1, s_2, s_3). The population mean, m, and the reciprocal effects (r_1, r_2, r_3) for individual crosses account for the remaining 1 and 3 *df* respectively. Reduction of the GCA parameters from 3 to 2 results from the standard subtraction of the final GCA column from the preceding two GCA columns. The $[X]$ matrix for Griffing's analysis, with its nine parameters corresponding to the nine observations, is thus:

Parameters for three-by-three diallel cross; Griffing's analysis (1.3)

Progenies:	m	g_1	g_2	s_1	s_2	s_3	r_1	r_2	r_3
y_{11}	1	2	0	−1	−1	0	0	0	0
y_{12}	1	1	1	1	0	0	1	0	0
y_{13}	1	0	−1	0	1	0	0	1	0
y_{21}	1	1	1	1	0	0	−1	0	0
y_{22}	1	0	2	−1	0	−1	0	0	0
y_{23}	1	−1	0	0	0	1	0	0	1
y_{31}	1	0	−1	0	1	0	0	−1	0
y_{32}	1	−1	0	0	0	1	0	0	−1
y_{33}	1	−2	−2	0	−1	−1	0	0	0

For Hayman's analysis, the nine parameter $[X]$ matrix, is:

Parameters for three-by-three diallel cross; Hayman's analysis (1.4)

Progenies:	m	j_1	j_2	j_{12}	j_{13}	j_{23}	r_1	r_2	r_3
y_{11}	1	2	0	−2	−1	0	0	0	0
y_{12}	1	1	1	1	1	1	1	0	0
y_{13}	1	0	−1	1	0	−1	0	1	0
y_{21}	1	1	1	1	1	1	−1	0	0
y_{22}	1	0	2	−2	0	−1	0	0	0
y_{23}	1	−1	0	1	−1	0	0	0	1
y_{31}	1	0	−1	1	0	−1	0	−1	0
y_{32}	1	−1	0	1	−1	0	0	0	−1
y_{33}	1	−2	−2	−2	1	1	0	0	0

The *df* for Hayman's additive (a's$=j_1, j_2$) and dominance (b's$=j_{12}, j_{13}, j_{23}$) effects are the same as for Griffing's GCA and SCA effects. If the number of model parameters equals the number of observations available, an exact rather than a least-squares solution results. This equality may occur because of deficiencies in the experimental data available, or where the $[X]$

Table 1. Diallel 1: F_1 observed and predicted values for mean seed weight (1,000 K weight in g) of parents and crosses for a six-parent diallel of flax cultivars. Predicted means from Griffing/Hayman models shown in brackets alongside each observed mean. Parameters in minimum adequate model shown at bottom

Female	Male parents						Means
	P	M	R	D	S	L	
P	5.24 (5.45)	5.23 (5.22)	5.39 (5.26)	5.42 (5.11)	4.79 (5.08)	4.82 (5.94)	5.15 (5.19)
M	4.73 (4.52)	4.36 (4.29)	4.61 (4.63)	3.93 (3.96)	3.79 (4.04)	4.15 (4.46)	4.26 (4.32)
R	7.62 (7.21)	6.49 (6.68)	6.98 (7.02)	6.53 (6.81)	6.10 (5.87)	6.62 (6.64)	6.72 (6.71)
D	7.16 (6.94)	6.76 (6.93)	7.01 (6.81)	7.23 (6.60)	6.69 (6.67)	6.03 (6.03)	6.81 (6.66)
S	5.24 (5.08)	4.95 (4.97)	5.88 (5.87)	5.08 (4.64)	5.07 (4.72)	4.81 (4.67)	5.17 (4.99)
L	5.09 (5.04)	4.66 (4.46)	4.89 (5.00)	5.07 (5.19)	4.79 (4.67)	4.62 (4.62)	4.85 (4.83)
Means	5.85 (5.71)	5.41 (5.43)	5.79 (5.77)	5.54 (5.39)	5.21 (5.18)	5.18 (5.24)	5.50 (5.45)

Parental mean = 5.59 (5.45), progeny mean = 5.48 (5.45)

F_1 parameters required in minimum adequate model using Griffing's model as basis

Analysis of variance for minimum adequate model for F_1

Total (uncorrected) sum of squares	105,150.14764	36 *df*
Regression sum of squares	105,126.44080	14 *df*
Residual sum of squares	23.70684	22 *df*

No.	Parameter no.	Estimate	Standard error	*t*-test
1	1	5.45034	0.03634	149.96
2	3 = GCA	−0.57990	0.03736	15.52
3	4 = GCA	0.78418	0.06247	12.55
4	5 = GCA	0.57607	0.05390	10.69
5	6 = GCA	−0.36663	0.04942	7.42
6	22 = recip.	0.34888	0.08761	3.98
7	23 = recip.	−0.97886	0.19957	4.90
8	24 = recip.	−0.91743	0.22017	4.17
9	27 = recip.	−1.02794	0.13834	7.43
10	28 = recip.	−1.48866	0.12773	11.65
11	29 = recip.	−0.46690	0.07697	6.06
12	33 = recip.	0.82086	0.18452	4.45
13	34 = recip.	1.01402	0.09921	10.22
14	35 = recip.	0.42113	0.05904	7.13

GCA, General combining ability effect; recip., reciprocal difference

matrix contains the most complex relevant model. From a programming viewpoint, it is simpler to reduce a complex initial model until adequate to explain the data, rather than the converse. By starting this way, parameters may be selectively and sequentially removed until the remaining model is just adequate, i.e. the model's residual sum of squares would become significant if one more parameter were removed. This approach provides the minimum adequate model for the data. Such reduction of the original model is feasible given data with appropriate weights and a suitable computer program. By ranking the *t*-test value for each parameter (i.e. [parameter-0]/standard error) and locating the minimum, the potentially redundant parameter corresponding to this minimum *t* can be eliminated from the model, which is then re fitted to the data. Elimination/refitting cycles continue until the minimum model is achieved (Tyson 1973). The model's parameters may already be partially or completely orthogonal, or may be made completely orthogonal in the first stages of computation, using, for example, the Gram-Schmidt procedure (Anton 1977) for obtaining ortho-normal bases.

Although matrix methods are computationally much heavier than equivalent scalar formulae for the same analyses, and may need large design [*X*] matrices, both problems are readily programmable. The outstanding advantages of matrix methods are the ability to modify the model fitted to the data and obtain weighted estimates of the parameters in the model, plus their standard errors. Computer programs (Microsoft compiled BASIC, Macintosh II) embodying all these facilities were written here.

Results

Diallel 1

Mean seed weights, averaged over the three replicates in the F_1 and six in the F_2 generation, are shown in Tables 1 and 2; their variances are at the bottom of Table 4. For

Table 2. Diallel 1: F_2 observed and predicted (Griffing/Hayman models) values for mean seed weight (1,000 K weight in g) of parents and crosses for a six-parent diallel of flax cultivars. Array variances, covariances in F_1, F_2 shown at bottom

Female	Observed means Male parents						Means
	P	M	R	D	S	L	
P	4.63	4.61	5.62	6.21	5.36	4.98	5.24
M	4.78	4.68	5.69	5.65	4.70	4.73	5.04
R	5.88	5.85	6.43	6.53	5.79	5.88	6.06
D	6.20	6.08	6.93	6.71	5.79	5.78	6.25
S	5.07	4.79	5.78	6.03	4.27	5.42	5.23
L	5.08	4.73	5.81	6.08	5.24	4.63	5.26
Means	5.27	5.12	6.04	6.20	5.19	5.24	5.51

Parental mean = 5.23; progeny mean = 5.57

Predicted means from Griffing model (predicted means from Hayman's model in brackets):

Female	Male parents						Means
	P	M	R	D	S	L	
P	4.72 (4.59)	4.82 (4.83)	5.73 (5.66)	6.06 (6.18)	5.04 (5.08)	5.01 (4.99)	5.23 (5.22)
M	4.82 (4.83)	4.68 (4.68)	5.64 (5.65)	5.59 (5.58)	4.81 (4.77)	4.78 (4.85)	5.05 (5.06)
R	6.02 (5.93)	5.64 (5.65)	6.35 (6.63)	6.87 (6.79)	5.86 (5.74)	5.83 (5.82)	6.10 (6.09)
D	6.06 (6.18)	6.06 (6.05)	6.87 (6.79)	6.72 (6.73)	6.04 (5.90)	5.87 (5.83)	6.27 (6.25)
S	5.04 (5.08)	4.81 (4.77)	5.86 (5.74)	6.04 (5.90)	4.26 (4.29)	5.00 (5.34)	5.17 (5.19)
L	5.01 (4.99)	4.78 (4.85)	5.83 (5.82)	6.16 (6.13)	5.00 (5.34)	4.63 (4.61)	5.24 (5.29)
Means	5.28 (5.27)	5.13 (5.14)	6.05 (6.05)	6.24 (6.22)	5.17 (5.19)	5.19 (5.24)	5.51 (5.52)

Parental mean = 5.23 (5.26); progeny mean = 5.57 (5.57)

Vr, Wr values for arrays in each generation

Parents	F_1 Vr	F_1 Wr	F_2 Vr	F_2 Wr	F_1 [Wr + Vr]	F_2 [Wr + Vr]
P	0.49775	0.83816	0.37874	0.56885	1.336	0.948
M	0.28515	0.60778	0.32632	0.59343	0.893	0.920
R	0.33426	0.65323	0.17735	0.43399	0.987	0.611
D	0.52891	0.86055	0.16084	0.40235	1.389	0.563
S	0.39823	0.75188	0.38720	0.56134	1.150	0.949
L	0.28119	0.63333	0.30743	0.48808	0.915	0.796
Wr/Vr slopes:	$b_1 = 1.011 \pm 0.06859$		$b_1 = 0.7311 \pm 0.16179$			

F_1 mean seed weight (Table 1), row marginal totals compared to corresponding column totals clearly indicate reciprocal differences. No correlation occurs between the 36 progeny means and variances in either generation. Extraction, for F_1 and F_2 data, of minimum adequate models (Griffing and Hayman) is followed by results from plotting array covariances against array variances.

The $[X]$ matrix for Griffing's model and the 36 F_1 or F_2 data contains 36 parameters comprising the overall mean, GCA effects, SCA effects and reciprocal effects. The resultant 36-by-36 $[X]$ matrix is not given here, but is a straightforward extension of that shown for a three-by-three diallel at (1.3). Parameters two–six deal with GCA effects, 7–21 deal with the 15 SCA effects, and the remaining 22–36 deal with the 15 individual reciprocal differences. To fit these parameters to the F_1 and F_2 data and extract the minimum adequate models, a computer program carrying out the weighted least-squares solution in Eq. (1.1) was used.

For GCA and SCA effects, the minimum adequate F_1 model is shown in Table 1. Fourteen of the original 36 parameters summarize the F_1 means, and are all significant. The analysis of variance is also shown; the residual sum of squares is not significant. Besides the mean, only parameters dealing with GCA and reciprocal effects occur; no SCA effects are present. Nine of the possible

Table 3. Diallel 1: F_2 parameters required in minimum adequate model using Griffing's model as basis

Total (uncorrected) sum of squares	88477.19519	36 *df*
Regression sum of squares	88454.15731	12 *df*
Residual sum of squares	23.03788	24 *df*

No.	Parameter no.	Estimate	Standard error	*t*-test
1	1	5.51634	0.02214	249.17
2	2 = GCA	−0.27128	0.03173	8.55
3	3 = GCA	−0.41787	0.02903	14.39
4	4 = GCA	0.55493	0.03197	17.36
5	5 = GCA	0.71631	0.04536	15.79
6	6 = GCA	−0.33152	0.03485	9.51
7	9 = SCA	0.22217	0.08068	2.75
8	10 = SCA	0.16317	0.04996	3.26
9	21 = SCA	0.40328	0.10001	4.03
10	23 = recip.	−0.13586	0.06469	2.10
11	28 = recip.	−0.23632	0.09106	2.59
12	35 = recip.	−0.14837	0.07496	1.98

Diallel 1: F_2 parameters required in minimum adequate model using Hayman's model as basis

Total (uncorrected) sum of squares	88477.19519	36 *df*
Regression sum of squares	88461.25157	15 *df*
Residual sum of squares	15.94361	24 *df*

No.	Parameter no.	Estimate	Standard error	*t*-test
1	1	5.50893	0.02070	266.19
2	2 = a	−0.25422	0.03059	8.31
3	3 = a	−0.41666	0.02863	14.55
4	4 = a	0.56136	0.03127	17.95
5	5 = a	0.74736	0.04202	17.78
6	6 = a	−0.33852	0.02962	11.43
7	7 = b_1	0.05652	0.00939	6.02
8	9 = b_2	−0.07230	0.01728	4.18
9	12 = b_2	0.01092	0.01824	3.90
10	14 = recip.	−0.14526	0.06468	2.24
11	19 = recip.	−0.23220	0.09113	2.55
12	26 = recip.	−0.14861	0.07496	1.98

GCA, General combining ability effect; SCA, specific combining ability effect; recip., reciprocal difference

15 reciprocal differences are significant, and only parent 1 (*P*) displays no general (GCA) combining ability. Predicted values from this model correspond closely ($r = 0.98$) to observed F_1 values (Table 1). The model's parameters clearly produce accurate predicted values for the F_1. The reciprocal effects show that large seeded parents produce large seeded progeny. While this is not unexpected, both this maternal effect and the occurrence of dominance towards higher mean seed weight seen in the F_2 have possible implications for survival of hybrid progeny in competition with the parental genotypes, from which they arose, and other co-existing genotypes.

Hayman's model for the six-by-six diallel initially contains 27 parameters; the ten omitted concern the non-

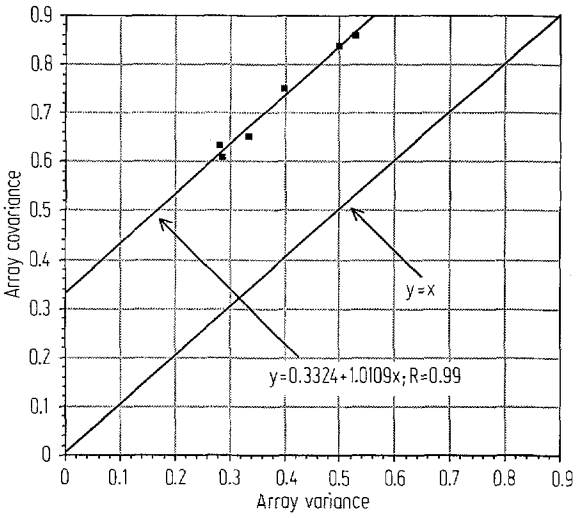

Fig. 1. F_1 array covariances compared to array variances

orthogonal b_3 items. The [*X*] matrix is not shown, but again is a straightforward extension of (1.4). In the F_1 data, 14 parameters are required in the minimum adequate model. These, and their analysis of variance, are identical to the Table 1 result. Parameters 22–35 in the Griffing-based model,dealing with reciprocal differences, correspond to parameters 13–26 in the Hayman-based model. Parameters three–six concern additive genetic effects and share an identity with Griffing's GCA parameters. No evidence of dominance occurs in the F_1, only additive genetic effects (excepting parent 1 = *P*, as before) and reciprocal differences are present. F_1 predicted values are thus identical to those of the Griffing minimum adequate model.

In the F_2 (mean seed weight data in Table 2, variances at the bottom of Table 4), the minimum adequate Griffing or Hayman-type models each need 12 parameters. These are shown with their standard errors in Table 3; all are significant, with non-significant residual terms in the analyses of variance. Here, the two models' predicted values (Table 2) are slightly different. Either model clearly predicts the F_2 data very closely; the omitted Hayman b_3 parameters are negligible.

In the F_2, the Griffing parameters show GCA effects for all parents, as do the Hayman parameters for additive genetic effects. However, SCA effects for three pairs of parents appear in the F_2. These pairs concern P and D, P and S, and S with L. The parameters (9, 10, 21) are all positive; the comparison in the [*X*] matrix is (progeny mean − parent mean). F_2 progeny for these pairs thus show dominance towards higher mean seed weights. The Hayman parameters show significant overall dominance in the departure of the progeny mean from the parental mean (param. $7 = b_1$), as well as dominance specifically for parents D and S (params. 9, 12, $= b_2$). The same three

Table 4. Diallel 1: Variances of observed means in F_1 and F_2. Mean seed weights for diallels 2 and 3 at bottom

F_1 Female	F_1 Male parents						Means
	P	M	R	D	S	L	
P	0.02370	0.00621	0.05114	0.11893	0.01791	0.07964	0.04959
M	0.17603	0.00681	0.01613	0.01623	0.04120	0.11621	0.06210
R	0.14981	0.18301	0.04454	0.12403	0.16963	0.03601	0.11784
D	0.08138	0.10210	0.04988	0.17643	0.00481	0.00063	0.06921
S	0.20988	0.00401	0.25000	0.16230	0.10431	0.16654	0.14951
L	0.03074	0.09524	0.22248	0.02058	0.00714	0.00351	0.06328
Means	0.11192	0.06623	0.10570	0.10308	0.05750	0.06709	0.08526

Parental mean = 0.05988; progeny mean = 0.09033

F_2 Female	F_2 Male parents						Means
	P	M	R	D	S	L	
P	0.01792	0.05355	0.00736	0.01463	0.06280	0.01510	0.02856
M	0.00708	0.00751	0.00626	0.02806	0.00867	0.01419	0.01196
R	0.00955	0.01128	0.01835	0.08730	0.01840	0.00928	0.02569
D	0.02109	0.01054	0.05128	0.09465	0.08553	0.01116	0.04571
S	0.00160	0.00642	0.01842	0.04154	0.01292	0.19718	0.04635
L	0.00542	0.02159	0.02242	0.01132	0.09887	0.01420	0.02897
Means	0.01044	0.01848	0.02068	0.04625	0.04787	0.04352	0.03121

Parental mean = 0.02759; progeny mean = 0.03193

Observed values for mean seed weight (1,000 K weight in g) in diallels 2 and 3

Female	Diallel 2 F_2 Male parents					Female	Diallel 3 F_1 Male parents				
	SC	RW	HO	A1	Means		R	M	A2	T	Means
SC	3.27	3.30	3.48	2.83	3.22	R	3.76	5.48	5.92	5.44	5.15
M	3.26	1.87	3.39	4.22	3.19	M	3.26	1.87	3.39	4.22	3.19
RW	3.82	4.52	4.14	3.56	4.01	A2	2.73	2.94	2.72	1.49	2.47
A1	2.94	3.30	3.24	2.29	2.94	T	3.45	5.02	2.09	4.93	3.87
Means	3.41	3.86	3.59	2.98		Means	3.30	3.83	3.53	4.02	

Parental mean = 3.39; progeny mean = 3.48

Parental mean = 3.32; progeny mean = 3.79

reciprocal differences are significant for either model in the F_2. This compares with the nine differences found in the F_1, so that maternal effects are clearly transient for mean seed weight with these parental genotypes and genotrophs. The pronounced maternal in the F_1 most likely obscure any dominance in the F_1, dominance which clearly surfaces in the F_2.

Array covariances (Wr) and variances (Vr) calculated from 1,000 K wt. data summed over all replicates for the F_1 and F_2 are shown at the bottom of Table 2 with the (Wr + Vr) values for each generation. The significant regression of F_1 array covariances (Wr) on corresponding array variances (Vr) is shown in Fig. 1. The slope is the expected value of 1.0; there is dominance, although the data regression position well above the line ($y = x$) passing through the origin indicates a very low dominance level, as would be expected from the Griffing/Hayman model results above.

The low level of dominance seen in the Wr/Vr plot for the F_1 is again a side-effect of the pronounced maternal effects present in the F_1 crosses. These side-effects are probably seen also in the plots of parental 1,000 K wt. against the estimates of the relative numbers of dominant genes in each parent, given by (Wr + Vr). The plot for the F_1 data is shown in Fig. 2, with the F_2 result shown in Fig. 4. These plots are expected to give significant linear regressions if dominant genes controlling seed weight act uniformly in the same direction, either towards increasing (negative regression slope), or decreasing (positive regression slope) seed weight. The F_1 regression is not

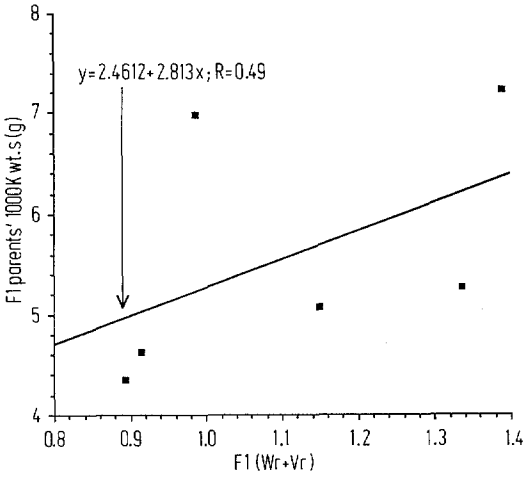

Fig. 2. F_1 parental 1,000 Kwt.s compared to (Wr + Vr)

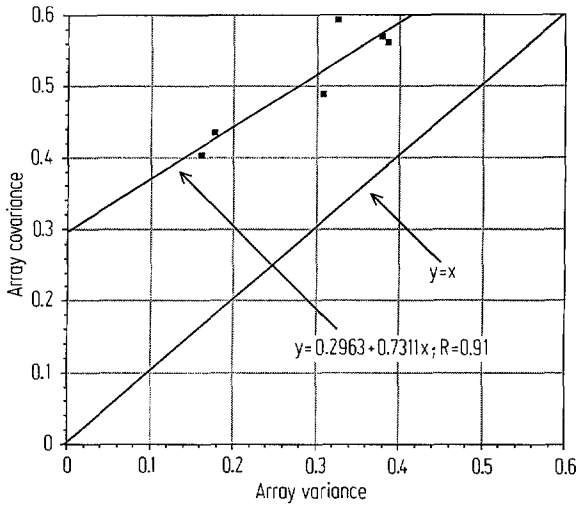

Fig. 3. F_2 array covariances compared to array variances

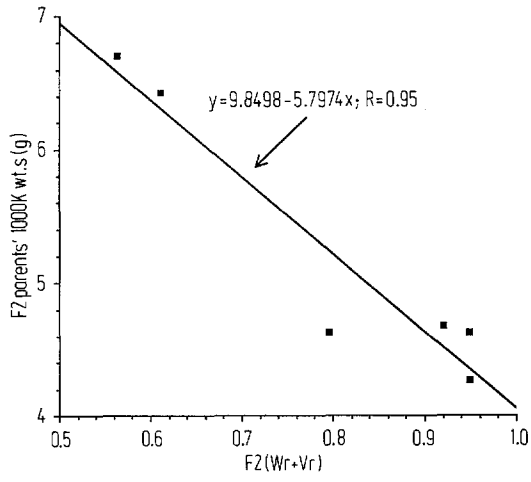

Fig. 4. F_2 parental 1,000 Kwt.s compared to (Wr + Vr)

significant. The highly significant, negative F_2 regression shows clearly that dominance is largely unidirectional towards higher seed weight.

In the F_2 Wr/Vr plot shown in Fig. 3, there is again a significant linear regression of Wr on Vr, but the overall position of this line relative to the line passing through the origin, i.e. $y = x$, shows that the level of F_2 dominance is slightly higher than that in the F_1 data of Fig. 1. This is corroborated by the result obtained from extracting the minimum adequate Griffing and Hayman models for the F_2 (Table 3); dominance surfaces in the F_2 as maternal effects disappear. The significant regression slope of 0.7311 does not depart significantly from the expected 1.0, and although there is a slight suggestion of possible disturbance through non-allelic interaction, it may be taken that the simplest genetical model omitting this is satisfactory for analysis purposes here.

The ordering of the six points along the best-fitting line (Fig. 3) is identical to that for the significant regression (Fig. 4) if F_2 parent seed weights on their corresponding (Wr + Vr) values, in which the inverse relationship signals unidirectional dominance towards higher mean seed weight. Figures 3 and 4 effectively summarize the genetic control of seed weight in this group of parental genotypes when transient maternal effects have disappeared. The F_3 and later generations could reasonably be expected to provide similar conclusions in the absence of interaction between genotypes of later generations (with lower heterozygosity) and any changes in environmental factors.

In Fig. 4 the equation for the regression line suggests that a hypothetical genotype containing all possible dominants for higher seed weight would have a 1,000 K wt. of nearly 10 g, on the basis of the intercept. This putative 1,000 K wt. coincides with the actual values for the largest seeded flax genotypes in the USDA World Flax Collection grown under optimal conditions.

There is a clear separation between the points for the oilseed (linseed) genotypes D and R, with low (Wr + Vr) and high 1,000 K wt. values, and the remaining four fibre type parents. The F_2 values for (Wr + Vr) thus provide a good indication of the relative numbers of dominant genes for mean seed weight in each of the six parents. The ranking is D > R > L > M > P > S, i.e. genotypes D (= Dakota) and R (= Royal) contain relatively most of the dominant genes for high mean seed weight. It is noteworthy that there is a difference between the S and L genotrophs in their (Wr + Vr) values; L has relatively more dominants for high seed weight than S. For the F_2 data, L shows a slightly higher mean seed than S, so that within the two genotrophs there may be a unidirectional dominance at seed weight loci. However, the inconsistent F_1 and F_2 values for S and L make this speculative. If there are indeed differences in the relative numbers of dominant seed weight genes in S compared to L, the way

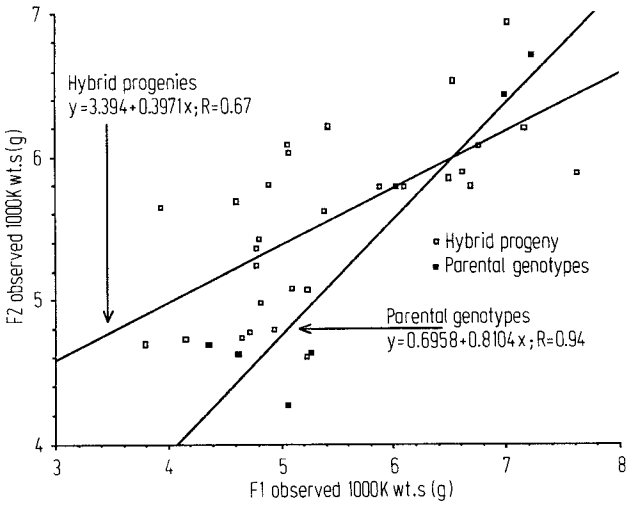

Fig. 5. F_2 compared to corresponding F_1 observed 1,000 Kwt.s

in which this might have originated through the environmentally induced heritable changes within the single Stormont Cirrus genotype is not clear.

The final point is the relation between observed F_1 results and corresponding observed F_2 results from the succeeding season. The actual correlation between the F_1 and F_2 observed values is plotted in Fig. 5, with the F_2 values on the *y* axis, and the points for the six parents and 30 hybrid progenies separated. The two successive generations grown in the field confounded genetic changes with seasonal differences. Individual regression lines for parents and their progenies are fitted and the difference between these lines, and thus the seasonal and F_1/F_2 relationships, is clear. The range of 1,000 K wt. among the six parents is approximately the same in the F_1 as in the F_2, but for the progenies it is somewhat less in the F_2. The parents' regression line suggests that between the two growing seasons, 1,000 K wt. was relatively stable. For the hybrids, the drop in heterozygosity, segregation in the F_2 and differences between growing seasons may all have contributed to the F_1 to F_2 changes seen in the plot.

Among the six parental points, the two highest 1,000 K wt. are the linseed types R and D; these occur at the top right of the parents' regression line. The four other parent points are the fibre types P and M, plus the L and S genotrophs derived from the fibre type Stormont Cirrus. The overall positions of the parent points relative to the progeny point – viewed from the F_1 axis – concur with the virtual identity between parent and F_1 means seen in Table 1. Viewed from the F_2 axis, the higher progeny mean relative to the parental mean (see Table 2) reflects the F_2/parent point positions, suggesting dominance in the F_2 towards higher 1,000 K wt. in at least some of the crosses.

Diallels 2 and 3

The mean seed weight data for diallels 2 and 3 are shown at the bottom of Table 4. There are two main points to be made with these data. The first is that they each include parental genotypes with strongly contrasting characteristics. The parental seed weights in Table 4 show the difference between the very small seeded Ethiopean types A1 and A2 and the oilseed types RW, HO and R. Wild type T (Turkish) represents a very large seeded genotype from a natural population. The second point is that the overall progeny mean is distinctly larger than the overall parental mean for each of the two diallels. There is one pronounced case of (over) dominance towards lower seed weight (A2 × T; diallel 3), but in the majority of individual crosses, the F_1 or F_2 hybrid means display dominance towards the larger seeded parent; in some cases over-dominance occurs.

Discussion

Although the origin of the tissues comprising the higher plant seed is complex with differences in ploidy, the mean seed weight character displays continuous variation and is best analysed with methods devised for quantitative inheritance. Fitting genetic models through matrix methods for weighted least-squares procedures has been combined here with the elimination of redundant parameters. This results in the most compact models accounting for the experimental variation in 1,000 K wt., simultaneously coping with the inevitable variation between replicates. Weighting is warranted by the significant, observable heterogeneity among the variances for cell means in the F_1 and especially the F_2. The degree of benefit provided by weighting, given increasing heterogeneity among variances, will be examined elsewhere as a separate topic, but two points can be made. Firstly, weighted estimation utilises all the experiment's information and will thus never be less 'useful' than an unweighted solution on the same data, even if $[V]$ in Eq. (1.1) approaches an identity matrix. Secondly, increasing heterogeneity of cell mean variances implies that weighting will provide results that reflect what is happening among those data which are more reliable and consistent in the experiment. For the F_2 here, this is particularly valuable. In the randomized block design, it implies also that marked genotype by environment interaction, causing some cells to be more variable over replicates than others, is de-emphasized by weighting, so that the results stress the behaviour of those genotypes which are, in general terms, more stable across varying environments.

The models extracted here describe genetic control over flax mean seed weight, but one may speculate on the potential effects of mean seed weight variation for the survival/fitness of genotypes in an inbreeding species

such as flax. Such effects have, for example, been of interest to plant breeders in a range of crop plants since the advent of modern plant breeding methods in the nineteenth century.

In examining the effect of variation in mean seed weight on the plants' subsequent performance, the sources of mean seed weight differences and the situations in which their effects might be tested have to be defined. A single, completely homozygous plant reproducing by 100% selfing can produce seed varying in mean weight as a result of timing/nutritional/positional effects during its flowering/seed setting phase. The individual seeds are, barring mutation or environmentally induced heritable changes (Durrant 1971), genetically identical. Any outcrossing could introduce genetic differences between individual seeds on the same plant with a potential impact on mean seed weights, effects confounded with the timing, etc., factors above. Micro-environmental differences between two genetically identical plants may also cause mean seed weight variation, again with the same comment regarding Durrant's findings. Between genetically identical plants, outcrossing may produce 1,000 K wt. differences. Lastly, genetic differences among the seed bearing plants are an obvious source of 1,000 K wt. differences.

The impact of mean seed weight differences on growth/performance for plants in populations of one species can be examined firstly where inter-plant competition is eliminated by isolation. This case is only complicated by the overall nutrient level available to all individual plants and interactions between their stage of development and variation in environmental conditions. This is a consideration in field studies, but not in growth chambers. Competition between plants of identical genotype, introducing the factor of plant density stress on top of the previous considerations, is relevant to agriculture with improved populations of single genotype inbreeders or F_1 hybrids. Finally, competition between plants of different genotypes encompasses more realistic situations for agriculture, either for an inbreeding cultivar assembled from genetically different lines, or an outbreeding cultivar created as a synthetic or progeny from chosen crosses. This last case (within species competition) is also relevant to natural populations. In many natural plant communities, a final stage involves competition between plants of different species.

Harper et al. (1970), in reviewing the limited literature on the genetics of mean seed weight control, note that selection, either natural or artificial, for increased/decreased mean seed weight is usually successful, since the character is generally associated with high heritability values. An inverse correlation between seed weight and seed numbers appears in experimental data; selection for increased seed weight/size results in decreased numbers of seed per plant. Selection for overall yield is most likely to cause the larger shift in seed number per plant rather than mean seed weight. Harper et al. quote results from a flax diallel indicating little dominance, as in the results here, where the progeny show dominance towards higher 1,000 K wt. which emerges in the F_2 after transient F_1 reciprocal differences have mostly vanished. Dominance towards larger mean seed weight is also seen in the two additional diallels here, supporting cautious generalization beyond the six chosen parents in the main data set analysed. Manner (1956) has also reported that mean seed weight in the F_2's of a majority of 27 different flax crosses showed dominance towards larger mean seed weight.

The effects of mean seed weight variation on the subsequent performance of plants of the fibre and oilseed genotypes used in this diallel have been investigated (Tyson, unpublished; data not shown here). Within each of the four genotypes, individual plants were genetically identical and completely homozygous as a result of the breeding system, experimental conditions and handling of successive generations. Within each of the four genotypes, mean seed weight variation was produced by collecting seed at different stages of ripeness and by placing one group under low light conditions during ripening. The control consisted of plants on which seed ripened normally under high fertility and full light intensity conditions. A range of seed weights was thus available within each genotype; between each seed weight group there were, in the absence of mutation or induced heritable changes, no genetic differences. The progeny plants from the four seed weight groups of each genotype were examined under conditions eliminating inter plant competition. Plant heights at the halfway stage in the life cycle showed changes due to 1,000 K wt. differences; at maturity, these changes had essentially vanished.

In the case of competition between plants of the same or different genotypes a partial answer is available in data (Tyson, again unpublished) concerning the effects of mean seed weight differences foa a wheat genotype. The final grain yield performance of plants produced from different seed weight classes under low fertility conditions with high competition among plants of the different seed weight classes clearly showed that mean seed weight was a significant factor, and that over one generation at these conditions the initial mean seed weight differences were erased, suggesting that the impact would be transient. The relationship between mean seed weight and subsequent grain yield was positive in the highly competitive conditions of this wheat experiment. This positive correlation appears in a number of studies by other workers in a variety of crop plants, (e.g. *Trifolium subterraneum* quoted by Harper et al. 1970), strengthening the viewpoint that larger mean seed weight/size confers advantages, albeit temporary, for successful completion of flowering and seed production for the next generation.

An unresolved question in plant reproductive strategy is the obvious success of many inbreeding species, including *Linum usitatissimum*, a virtually completely inbreeding species occurring across the north temperate/middle east zones, and showing 0%–3% outcrossing in experimental populations in the field (Dillman 1946). Genetic variability has to be maintained within any one such natural population in the face of continuous selfing, with selection eliminating some of the more or less homozygous genotypes. Allard et al. (1968) have suggested that sporadic, limited outbreeding may be the source of heritable variation in some species and populations examined, but concede that the observed variation cannot always be reconciled with the outbreeding apparently occurring. In some inbreeding populations heterozygotes, which may arise through such events, persist in the population at frequencies far higher than expected from the return to homozygosity under complete selfing, presumably due to higher heterozygote fitness.

If higher mean seed weight could at least ensure seedling survival and higher mean seed weight results from dominant genes, a fortuitous, uncommon outcross between two or more parental genotypes differing in mean seed weight would generate a proportion of F_1 progeny which are large-seeded because of maternal effects, and in the F_2 and later generations contain mostly large-seeded individuals due to dominance. Thus, progeny from a sporadic occurrence of outbreeding should survive beyond the seedling stage to allow a first generation screening with comparison, through response to environmental conditions and competition, to both parental and other genotype neighbours. Survival during the F_2 exposes new segregants for other characteristics to natural selection. Mechanisms aiding survival of hybrid progeny should presumably be advantageous, in that potentially valuable new genotypes from infrequent outcrosses will not be lost immediately from the inbreeding population. A combination of mean seed weight effect and genetic control of mean seed weight may underlie the survival of large seeded genotypes in natural conditions, the more so if seed size and seed number per plant are inversely correlated. It should not be difficult to check the possibilities, to find whether dominance towards larger mean seed weight coupled with a competitive advantage for larger mean seed weight are common amongst a wide range of inbreeding species.

Acknowledgements. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, to whom thanks are extended.

References

Allard RW, Jain SK, Workman PL (1968) The genetics of inbreeding populations. Adv Genet 14:55–131

Anton H (1977) Elementary linear algebra. Wiley, New York

Bartlett MS (1937) Some examples of statistical methods of research in agriculture and applied biology. J R Stat Soc Suppl 4:137–183

Cavalli LL (1952) An analysis of linkage in quantitative inheritance. In: Reeve ECR, Waddington CH (eds) Quantitative inheritance. Her Majesty's Stationary Office, London

Dillman AC (1946) Classification of flax varieties. USDA Techn Bull No. 1054

Draper NR, Smith H (1968) Applied regression analysis. Wiley, New York

Durrant A (1965) Analysis of reciprocal differences in diallel crosses. Heredity 20:573–607

Durrant A (1968) Phenotype analysis of diallel crosses. Heredity 24:541–560

Durrant A (1971) Induction and growth of flax genotrophs. Heredity 27:277–298

Fieldes MA, Tyson H (1984) Possible post-translational modification and its genetic control, in flax genotroph isozymes. Biochem Genet 21:99–114

Griffing B (1956) Concept of general and specific combining ability in relation to diallel crossing systems. Aust J Biol Sci 9:463–493

Harper JL, Lovell PH, Moore KG (1970) The shapes and sizes of seeds. Ann Rev Ecol Syst 1:327–356

Harvey WR (1960) Least squares analysis of data with unequal sub-class numbers. Bull ARS-20-8, USDA

Hayman BI (1954 a) The analysis of variance of diallel tables. Biometrics 10:235–244

Hayman BI (1954 b) The theory and analysis of diallel tables. Genetics 39:789–809

Jeswani LM, Murty BR (1963) Components of association for some yield factors in linseed. Indian J Genet 25:163–175

Khan MA (1963) Genetic analysis of varietal differences within *Linum usitatissimum* (flax and linseed). PhD Thesis, University of Wales

Manner R (1956) Studies on seed-setting and seed yield in oil flax. Meddelande Från Gullåkers Växtförädlingsanstalt, Hammenhög Nr. 12. Paper presented at Faculty of Agriculture and Forestry, University of Helsinki, Finland

Mather K, Jinks JL (1982) Biometrical Genetics, 3rd. ed. Chapman and Hall, London

Morley-Jones R (1965) Analysis of variance of the half diallel table. Heredity 20:117–121

Smith WE, Fitzsimmons JE (1964) Maternal inheritance of seed weight in flax (*Linum usitatissimum*). Can J Genet Cytol 6:244

Tyson H (1973) Cytoplasmic effects on plant weight in crosses between flax genotypes and genotrophs. Heredity 30:327–340