Theor. Appl. Genet. 53, 181-190 (1978)

© by Springer-Verlag 1978

Investigations of the Seed Protein Content of Several Pea Genotypes Grown in Two Different Years

H.D. Quednau and G. Wolff
Institute of Genetics, University of Bonn (Federal Republic of Germany)

Summary. Seventeen X-ray and neutron induced mutants of the commercial variety 'Dippes gelbe Victoria' were analyzed with regard to their seed protein percentage. The interaction of genotypic and year effects in 1975 (normal weather conditions) and 1976 (extremely hot and dry) was also taken into consideration. To avoid undiscoverable environmental bias, the plants were grown in a non-standard three-dimensional layout. Biometric analysis was done by using the theory of the general linear model with a formula-processing computer program. In the first year, significant genetically caused differences were found in the material. The bifurcated mutant 157A was especially of considerable interest because an improved protein content was combined with relative good yield. In the second year, no significant differences between the mutants were revealed, but all genotypes showed a similar good protein value of about 27%.

Key words: *Pisum sativum* – Seed protein content – Mutation – Genotype-year-interaction

Introduction

Since the beginning of the sixties, the vital role of protein in world nutrition has become apparent, and consequently, an intense research work has begun all over the world with the purpose of improving protein quality and/or quantity by means of mutation and selection in crops, vegetables and pulses.

In general, the first step in finding protein mutants is by examining mutant populations on the basis of phenotypical characters. An appropriate selection has to follow the detection, taking into consideration all interacting plant characters such as yield, kernel size, flowering time, and other morphological and physiological characters in addition to the interacting environmental factors that mask the true genetic variation.

In 1964, Mertz et al. published the most interesting findings on opaque-2 maize. The main favourable characteristic of this mutant is the improvement of the lysine content in seed protein caused by a changed proportion of protein sub-fractions. The induction of the protein improved mutant in barley (1508) (Doll et al. 1974) and sorghum (Singh and Axtell 1973) was a further step in this direction.

These improvements in maize, barley and sorghum are due to qualitative characters. These are more easily deliminated as they show a discontinuous distribution and are less influenced by environmental factors. Quantitative characters on the other hand – for example, the quantitative improvement of seed protein content – show continuous variations and often exhibit small differences between the mutant lines. Added to this, they are highly susceptible to environmental factors.

A perusal of the literature indicates that most of the socalled 'protein mutants' show quantitative characters and while a few may be real 'protein mutants', the majority are simply modifications (cf. Proc. of IAEA 1969-1978). In those cases, the improved protein content is not genetically conditioned, and thus, the difference is not significant from the corresponding initial lines.

Since 1969, investigations in our institute on seed protein make-up of pea mutants have been carried out (Gottschalk and Müller 1974; Gottschalk and Müller 1970; Müller and Gottschalk 1973; Gottschalk et al. 1975; Wolff 1975; Gottschalk et al. 1976; Quednau and Wolff 1978). One hundred and thirty mutants were examined for their seed protein content. Most of these showed little or no deviations from the values of the initial line. But apart from this, mutants with clear positive as well as negative differences in seed protein were found. For the mutants exhibiting marked differences (up to 30 or 40% protein) there were no problems about their evaluation. However, in general, these mutants are of no practical and theoretical use as the high protein value is connected to an ex-

0040-5752/78/0053/0179/$ 2.00

treme reduction in yield (Wolff 1975) and is not combined with any shifting of protein-subfractions or of amino acid composition (Wolff unpublished). This is why 'mutants' which show slight deviations are of interest. A genetically fixed slight improvement of protein content, combined with an improvement in yield, or at least a similar yield as the initial line, provides a good base for an agriculturally usable line. The difficulty is – as has already been mentioned – to find these 'real mutants' and to deliminate them from modifications of the initial line. The other problem of interest is the effect of different years on genetically conditioned seed protein alterations. We grew our material for two cultivation periods (1975 and 1976) which were rather different concerning temperature and humidity.

Material and Methods

Our investigations were carried out on 15 X-ray- and neutron-induced mutants and 2 recombinants homozygous for several mutant genes and their initial line. All genotypes arose from the commercial variety 'Dippes gelbe Victoria' and belong to the mutant collection of our institute.

1. The fasciated genotypes

489C apical part of the stem strongly fasciated, all the flowers and pods clustered at the top of the plant
2797 morphologically identical to 489C
251A apical part of the stem linearly fasciated, flowers and pods distributed over a relatively long part of the stem
123 phenotypically intermediate between 489C and 251A
12C short internodes
176X narrow leaves, late flowering; recombinant arisen by a spontaneous crossing between mutant 489C and 176A (narrow leaves)

The genes for 'fasciation' of all theses mutants belong to a series of multiple alleles (Gottschalk 1977).

2. The dichotomously branched genotypes

1201A reduced penetrance, homozygous for *bif-1*
239CH similar to 157A, flower anomaly, phenotypical identical to 1201A
157A reduced penetrance, homozygous for *bif-2*: *bif-1* and *bif-2* are two non-linked polimeric genes
46C early flowering and ripening, reduced penetrance, homozygous for genes *efr* and *bif-1*

3. Mutants showing a favourable protein situation in former years

15 flower mutant with extremely low fertility
189 small greenish flowers, strongly reduced fertility
419D narrow leaves
488B abnormal leaves: yellow margins at the leaves and stipules

4. Mutants showing a negative protein situation in former years

105 chlorophyll mutant 'chlorina', internode slightly shortened, late flowering
486 very slight chlorophyll deficiency, 'chlorescens'
2314 margins of the leavelets undulated

Cultivation of the Experimental Material

From most of the genotypes under investigation, only a small number of seeds were available. Moreover, the environmental influence on the protein content is so strong that comparisons between plants grown in a distance of some meters are almost impossible. Therefore, we worked out an experimental design which minimized an uncontrollable environmental bias and gave sufficiently short confidence intervals for genotypic effects in spite of the relative small amount of seed material used. The plants were grown in Bonn in 1975 and 1976. The seeds were allocated to 4 groups: group 1, comprising the bifurcated genotypes plus the initial line; group 2, the fasciated genotypes plus the initial line; groups 3 and 4, those mutants showing an improved or decreased seed protein content in former years, plus the initial line. The material was cultivated in 5 rows; each row contained one set of these groups. From each genotype 10 seeds were sown with a distance of 2 cm between the single plants so that each group extended over a length of between 80 cm (groups 4) and 140 cm (group 1). In 1975, seeds of all genotypes were harvested but in 1976 plants of mutants 15 and 189 did not grow as the seeds were eaten by birds before germination.

Protein estimation

The total nitrogen content of the seeds on the basis of dry weight was determined by the Kjeldahl method. Each sample contained the seeds of one pod. The protein content was calculated by multiplying the nitrogen value by 6.25. From each genotype of each group we entered 5 replications into the Kjeldahl analysis.

Biometric model

The biometric analysis was based on the following model: The protein content P of the seeds of pod no. 1 from the genotype i and row j in year k (k = 1 for 1975 and k = 2 for 1976) can be expressed by a multiplicative model of the form

$$P_{i,j,k,l} = NP_{i,j,k,\,1} \cdot (C \cdot G_i \cdot W_{i,k} \cdot R_{j,k} \cdot E_{i,j,k,l})$$

where NP means the non-protein content of the seeds' dry weight, C is a reference value (expectation of the initial line in row 1, year 1), G_i is the genotypic effect of mutant i in year 1, $R_{j,k}$ is the environmental effect of row j in year k, which adds to the environmental conditions of the reference, $W_{i,k}$ is the interaction of genotype i with year k, and $E_{i,j,k,l}$ is the remaining error term. The genotypic effect in year 2 is thus expressed by $G_i \cdot W_{i,2}$. Before entering the analysis, the raw data were transformed by the transformation

$$Y_{i,j,k,l} = \ln\left(\frac{P_{i,j,k,l}}{NP_{i,j,k,l}}\right)$$

to get an additive model of a variable which was supposed to be normally distributed with a common variance:

$$Y_{i,j,k,l} = \ln(C) + \ln(G_i) + \ln(W_{i,k}) + \ln(R_{j,k}) + \ln(E_{i,j,k,l})$$
$$= c + g_i + w_{i,k} + r_{j,k} + e_{i,j,k,l} \sim N(c + g_i + w_{i,k} + r_{j,k}, \sigma^2)$$

So we get a three dimensional unbalanced factorial design which cannot be analysed by standard ANOVA methods. Therefore, the biometric evaluation was done using the theory of the general linear model.

Let m_1 and m_2 be the number of mutants which have been entered into the group under consideration in year 1 and 2, respectively. As we had 5 rows in both years and five replications within each elementary unit (i,j,k), the data vector $\vec{y}$ had been expected to contain 25 $(m_1 + m_2)$ pieces of data. Because of single outliers and some missing data in groups 3 and 4, the length of $\vec{y}$ was

Table 1. Part of the design matrix for the transformed protein data of group 1

	c	g_2	g_3	g_4	g_5	$w_{2,2}$	$w_{3,2}$	$w_{4,2}$	$w_{5,2}$	$r_{2,1}$	$r_{3,1}$	$r_{4,1}$	$r_{5,1}$	$r_{1,2}$	$r_{2,2}$	$r_{3,2}$	$r_{4,2}$	$r_{5,2}$
$Y_{1,1,1,s}$ (IL, row 1, year 1)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$Y_{1,2,2,s}$ (IL, row 2, year 2)	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
$Y_{2,1,1,s}$ (M 46C, row 1, year 1)	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
$Y_{2,3,1,s}$ (M 46C, row 3, year 1)	1	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
$Y_{3,4,2,s}$ (M 157A, row 4, year 2)	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0

Index s meaning an arbitrary replication number

reduced. The vector of expectation values is a linear combination of a parameter vector $\vec{p}$ with $10 + m_1 + m_2$ elements:

$$\vec{p} = (c, g_2, ..., g_{m+1}, w_{2,2}, ..., w_{m_2+1,2}, r_{2,1}, ..., r_{5,1}, r_{1,2}, ..., r_{5,2})^T.$$

We now construct a design matrix D such that $\vec{y} = D\vec{p} + \vec{e}$, where $\vec{e}$ is the vector of error terms.

Table 1 shows some lines of the design matrix of group 1 ($m_1 = m_2 = 4$) to illustrate the way by which it is built up.

A least-squares estimate of $\vec{p}$ is got by $\hat{\vec{p}} = (D^TD)^{-1} D^T \cdot \vec{y}$, and the vector of residuals $\vec{r}$ is calculated by $\vec{r} = \vec{y} - D \cdot \hat{\vec{p}}$. After computing $\hat{\vec{p}}$ and $\vec{r}$ we tested if the model was adequate for our data. Provided the model is correct, the distribution of $\vec{r}$ is a multivariate normal with expectation $\vec{o}$ and variance-covariance matrix $Q = \sigma^2 [I - D(D^TD)^{-1} D^T]$. Let C now be an orthonormal basis of Q. Then we get:

$$\vec{y} \sim N(D \cdot \vec{p}, \sigma^2) \Rightarrow C \cdot \vec{r} \sim N(\vec{o}, \sigma^2 \cdot I),$$

the dimension of the second normal distribution being the length of $\vec{y}$ minus the length of $\vec{p}$ (Putter 1967). In our test program we put $C = V^{-1/2}M$, where M contains the eigenvectors of Q and the diagonal matrix V, the corresponding eigenvalues. The resulting vector $C \cdot \vec{r}$ was tested for normality using the method of Gebhardt (1966), which is based on the 3^{th} and 4^{th} moments.

This test for global validity of our model needs a large amount of computer central storage and computer CPU-time. Therefore, it was performed only with the data of group 1, which turned out to be clearly unsignificant. Besides this we examined each group for the occurrence of outliers using a test of Lund (1975), which is based on critical values for the studentized residuals. When the model has proved to be adequate and no outliers are present, we can get an efficient estimate for σ^2 by: $\hat{\sigma}^2 = \vec{r}^T\vec{r}$ / (length $(\vec{y})$ − length $(\vec{p})$).

In the unbalanced design that we used, the estimates of (g_2, ..., g_{m+1}) and ($w_{2,2}$, ..., $w_{m+1,2}$) are not independent. As we wanted to investigate both parameter sets simultaneously, we had to set up the null hypothesis: H_O: $g_{i_1} = 0$ and $w_{i_2,2} = 0$ for $i_s = 2, ..., m_s + 1$. This hypothesis can be tested in the following manner: let H be a matrix with $m_1 + m_2$ rows and $(10 + m_1 + m_2)$ columns. Let in row r of H that element be 1 that corresponds to g_r (if $r \leq m_1$) or to $w_{r-m_1,2}$ (if $m_1 < r \leq m_1 + m_2$). Then we can reformulate the null hypothesis as follows:

$$H_O: H \cdot \vec{p} = \vec{o}.$$

If H_O is true, then

$$\frac{1}{\hat{\sigma}^2 \cdot df1} (H \cdot \hat{\vec{p}})^T [H(D^TD)^{-1} H^T]^{-1} (H \cdot \hat{\vec{p}}) \sim F_{df1,df2}$$

with df1 = number of rows of H = $m_1 + m_2$ and

df2 = length $(\vec{y})$ − length $(\vec{p})$ = 24 $(m_1 + m_2)$ − 10 − φ,

where φ is the number of missing values and outliers.

So with probability (1-α), the vector $\vec{q} = H \cdot \vec{p}$, which is a linear combination of the true parameters, is within the hyperellipsoid given by

$$(H \cdot \hat{\vec{p}} - \vec{q})^T [H(D^TD)^{-1} H^T]^{-1} (H \cdot \hat{\vec{p}} - \vec{q}) \leq \sigma^2 \cdot df1 \cdot F_{1-\alpha} (df1, df2) \quad [1]$$

In our case, the vector $\vec{q}$ consists of those elements of $\vec{p}$ which are entered into the null hypothesis, so we get

$$\vec{q} = (g_2, ..., g_{m_1+1}, w_{2,2}, ..., w_{m_2+1,2}).$$

Using inequation [1], we can first construct simultaneous confidence intervals for any number of linear contrasts of the q_i, and secondly test any subsets of the null hypothesis.

1) Simultaneous confidence intervals for linear contrasts $\vec{c}_s^T\vec{q}$ are achieved by optimizing $\vec{c}_s^T\vec{q}$ for $\vec{q}$ remaining on the surface of the confidence hyperellipsoid given by [1]. The restricted extreme values of $\vec{c}_s^T\vec{q}$ are

$$\min, \max (\vec{c}_s^T\vec{q}) = \vec{c}_s^T\hat{\vec{q}} \pm \hat{\sigma} \sqrt{df1 \cdot F_{1-\alpha}(df1, df2) \cdot \vec{c}_s^T Q \vec{c}_s}$$

$\hat{\vec{q}}$ | [1]

with $Q = [H(D^TD)^{-1} H^T]^{-1}$ and $\hat{\vec{q}} = H \cdot \hat{\vec{p}}$ (Scheffé 1953)

2) Let $C_s \cdot \vec{q}$ be a set of mutually independent contrasts, for which we want to know if any element of the confidence hyperellipsoid fulfils $C_s \cdot \vec{q} = \vec{O}$. In practice, the most important vectors formed by $C_s \cdot \vec{q}$ are $(g_2, ..., g_{m_1+1})^T$, $(w_{2,2}, ..., w_{m_2+1,2})^T$, and $(g_2 + w_{2,2}, ..., g_{m_2+1} + w_{m_2+1,2})^T$. In order to solve this problem, we search the minimum of $(\vec{q} - \vec{\bar{q}})^T Q^{-1} (\vec{q} - \vec{\bar{q}})$ with the restriction $C_s \cdot \vec{\bar{q}} = \vec{O}$. The method of Lagrange multipliers leads to:

$$1(\vec{q}, \vec{\lambda}) = (\hat{\vec{q}}-\tilde{\vec{q}})^T Q^{-1} (\hat{\vec{q}}-\tilde{\vec{q}}) + \sum_{j=1}^{z} (\lambda_j \cdot \sum_{k=1}^{w} (C_s)_{j,k} \cdot \tilde{q}_k)$$

$$\text{with } \frac{\partial 1}{\partial \tilde{q}_i} = 2 \cdot \sum_{t=1}^{w} Q^{-1}{}_{i,t} (\tilde{q}_t - \hat{q}_t) \sum_{j=1}^{z} \lambda_j (C_s)_{j,i}$$

$$\text{and } \frac{\partial 1}{\partial \lambda_j} = \sum_{k=1}^{w} (C_s)_{j,k} \cdot \tilde{q}_k$$

where w = length $(\vec{q})$ and z = number of rows of C_s.

Setting grad $1 = \vec{O}$ we get a system of linear equations with a symmetric coefficient matrix:

$$\left(\begin{array}{c|c} 2 \cdot Q^{-1} & C_s^T \\ \hline C_s & O \end{array}\right) \cdot \left(\frac{q^*}{\lambda^*}\right) = \left(\frac{2 \cdot Q^{-1} \cdot \hat{\vec{q}}}{\vec{O}}\right)$$

$$\text{If } (\hat{\vec{q}}-\vec{q}^*)^T \cdot Q (\hat{\vec{q}}-\vec{q}^*) = \hat{\sigma}^2 \cdot df1 \cdot F_{1-\beta}(df1, df2)$$

then there does not exist a vector $\vec{q}^*$ which fulfils $C_s \cdot \vec{q}^* = 0$, within a confidence-interval based on an error probability $< \beta$.

$$\text{So, if } \frac{(\hat{\vec{q}}-\vec{q}^*)^T Q (\hat{\vec{q}}-\vec{q}^*)}{\hat{\sigma}^2 \cdot df1} > F_{1-\alpha}(df1, df2)$$

we reject the hypothesis: $C_j \cdot \vec{q}^* = \vec{O}$, without increasing the error probability α.

The calculations were performed on the IBM 370/168 of the computing centre at the University of Bonn. For constructing the design matrix we used a symbol processing program (Quednau 1978) based on the computer language PL/I-FORMAC (IBM-Corp 1967). The numerical evaluation of the analysis followed Wottawa (1974).

The biometric model only was applicable to protein percentage. Yielding values were given as average values without any statistical calculations.

Results

The protein content of 18 genotypes was analysed by the Kjeldahl method. These genotypes were chosen under different aspects: genotypes of group 1 (4 bifurcated genotypes) and of group 2 (6 fasciated genotypes) were of some importance because of the good yield of most of them. This favourable character makes them interesting for plant breeding. An additional improvement in protein content would be of an extraordinary importance; in such a case the nutritional value of the plant would be increased. But if – on the other hand – the protein content of the seeds remains unchanged, the protein production per plant is nevertheless improved because of the higher yield. Genotypes of groups 3 and 4 were chosen because in former investigations we found that clear differences between them and the initial line exist.

The material was cultivated in two vegetation periods. In 1975, we had a normal spring and summer; that is, plants had sufficient rain and temperatures were in the range of 18 to 25° C. In 1976, the weather was completely unusual for our region: temperatures exceeding 30° C and no rain over a long period. We expected these two different vegetation periods to have influenced the seed protein content and lead to diverse values.

In Table 2, results of the tests for genotypic effects in both the years and the three subtests for genotypic effects in year 1975 and year 1976 separately, as well as for genotype-year-interaction are indicated. Figures 1-4 show the 5%-confidence intervals for the non-logarithmic genotypic effects G_i and $G_i \cdot W_{i,2}$. Within the intervals the value of exp $(\hat{g}_i)$ or exp $(\hat{g}_i + \hat{w}_{i,2})$, respectively is indicated. The genotype symbols are marked by +, ++, +++ if the confidence interval for the respective G_i with error probability $\alpha = 0.05$, 0.01 or 0.001 did not contain 1, so that a significant deviance of the mutant from the initial line in year 1975 has been proved. If for two mutants i and j of the same group the confidence interval for G_i/G_j did not contain 1, the mutant symbols are linked by a line which is marked by the respective cross symbol. This means that mutants i and j differ significantly from each other in year 1975.

For year 1976, no significant confidence intervals were found.

The practical meaning of a parameter A_i, which may be G_i or $G_i \cdot W_{i,2}$, is as follows: a pod of mutant i is expected to produce A_i-times the seed protein quantity of

Table 2. Results of tests of hypotheses

	Group 1	Group 2	Group 3	Group 4
d.f.	8,232	12,326	6,175	6,183
F-value for H_O: $\vec{g} = \vec{o}$. $\vec{w} = \vec{o}$	8.91^{+++}	8.27^{+++}	11.78^{+++}	17.41^{+++}
F-value for H_O': $\vec{g} = \vec{o}$	6.78^{+++}	6.80^{+++}	10.77^{+++}	17.28^{+++}
F-value for H_O'': $\vec{g} + \vec{w} = \vec{o}$	2.12^{+}	1.46^{-}	1.01^{-}	0.13^{-}
F-value for H_O''': $\vec{w} = \vec{o}$	2.56^{+}	1.40^{-}	1.11^{-}	8.21^{+++}

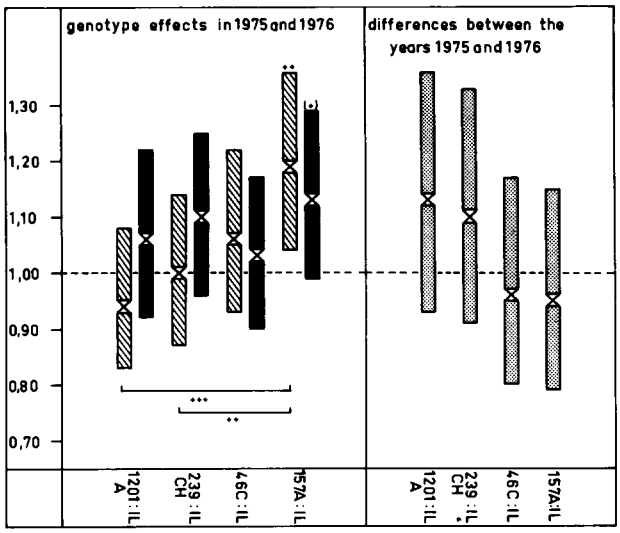

Fig. 1. 5%-confidence intervals for the non-logarithmic genotypic effects and genotype-year interactions, group 1 (bifurcated mutants). Light columns correspond to 1975, dark columns to 1976. Significant deviances from initial line are marked by crosses, significant differences between mutants by lines linking the appropriate symbols

a pod of the initial line with the same non-protein content of the seeds grown under the same environmental conditions. When we want to express the ratio R of the protein percentage of mutant i and initial line by the parameter A_i, we make use of the formula:

$$R_i = \frac{P_i}{P_i + NP} : \frac{P_{IL}}{P_{IL} + NP} = \frac{A_i \cdot P_{IL}}{A_i \cdot P_{IL} + NP} \cdot \frac{P_{IL}+NP}{P_{IL}}$$

$$= \left[\frac{1}{A_i} \cdot \left(1 - \frac{Q}{100}\right) + \frac{Q}{100}\right]^{-1},$$

where Q means the protein percentage of the initial line at a given environment. If for instance $A_i = 1.2$, then for Q = 20% we have R = 1.15 and for Q = 30% we have R = 1.13. From $A_i = 1$ of course results $R_i = 1$, independent from Q.

Considering all genotypes, 4 mutants with significantly improved protein percentage were found in 1975 (Figs. 1-4, light columns): mutant 157A (Fig. 1) and mutants 15, 189, and 419 (Fig. 3). For the values for 1976 (Figs. 1-4, dark columns) (seeds of mutants 15 and 189 were eaten by birds before germination) these differences between mutated genotype and initial line could not be found, though in mutant 157A the value is only slightly below significance (Fig. 1, dark column) and the subtest for genotypic effects in 1976 is significant for group 1 (Table 2, third line).

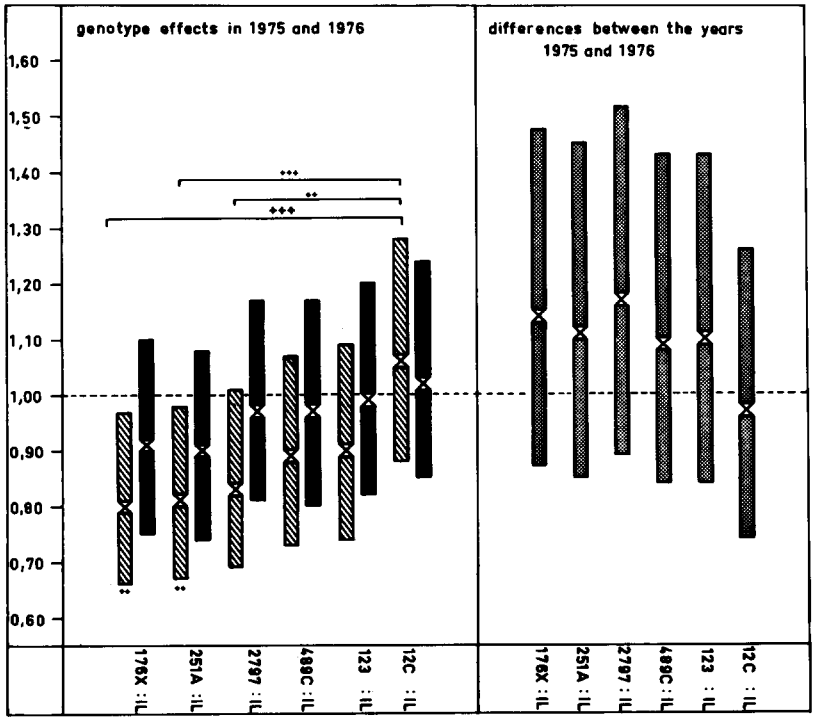

Fig. 2. 5%-confidence intervals for the non-logarithmic genotypic effects and genotype-year interactions, group 2 (fasciated mutants). Light columns correspond to 1975, dark columns to 1976. Significant deviances from initial line are marked by crosses, significant differences between mutants by lines linking the appropriate symbols

A significant negative deviation from the values of the initial line is found in genotypes 486 (Fig. 4) and 176X and 251A (Fig. 2) for the year 1975; in the second year these genotypes behave as the initial line did. All the other genotypes show no significant differences from the initial line neither in 1975 nor in 1976.

An interaction of genotypic and year effects was found in groups 1 and 4 (Table 2, last line). Within group 4 the high value of the interaction is mostly due to mutant 486 which behaves significantly different in the two years (Fig. 4, right part). As for group 1, we cannot say which of the mutants is responsible for the significant interaction effect.

To clear up the results presented in figures 1-4, the protein values (% protein per seed meal) of all genotypes are summarized in Fig. 5. The last 4 mutants (157, 419, 15, 189) of the left group and the first three genotypes (176X, 251A, 486) of the right group are significantly different from the initial line in 1975. The genotypes with improved protein content exhibit values which are about 3-4% higher in protein content (28.6, 28.8, 28.9, 29.2) than the initial line (25.5). The negative protein values (21.6, 21.7, 22.5) show the same range: 3-4% reduction in protein content. In 1976 no differences between the initial line and the mutated genotypes could be proven, as has already been mentioned. Protein values were more

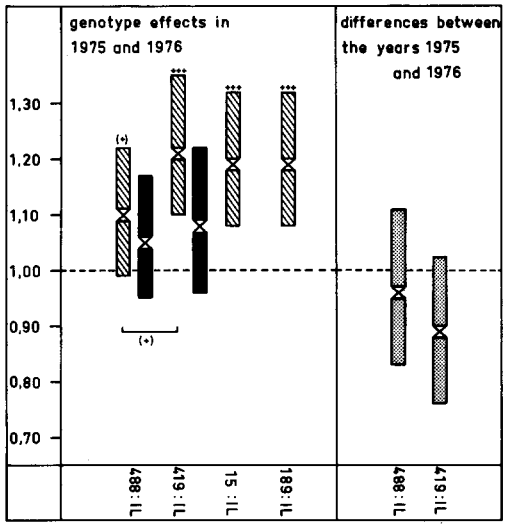

Fig. 3. 5%-confidence intervals for the non-logarithmic genotypic effects and genotype-year interactions, group 3.
Light columns correspond to 1975, dark columns to 1976. Significant deviances from initial line are marked by crosses, significant differences between mutants by lines linking the appropriate symbols

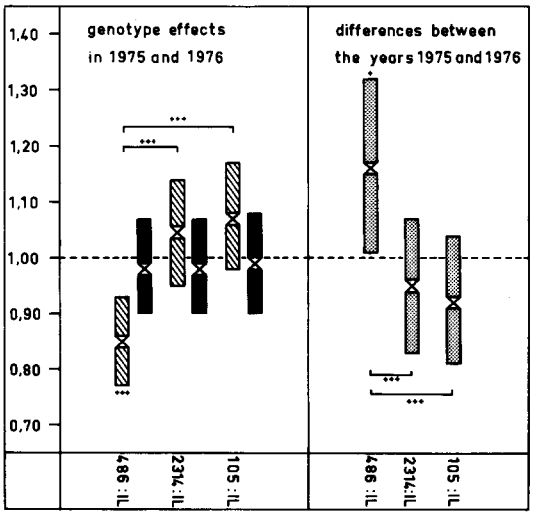

Fig. 4. 5%-confidence intervals for the non-logarithmic genotypic effects and genotype-year interactions, group 4.
Light columns correspond to 1975, dark columns to 1976. Significant deviances from initial line are marked by crosses, significant differences between mutants by lines linking the appropriate symbols

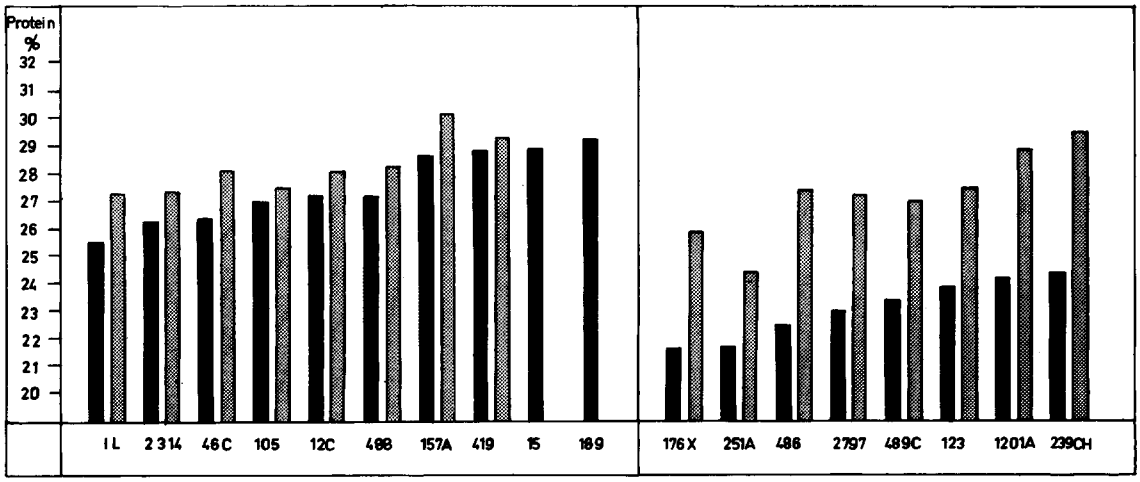

Fig. 5. Protein content per seed meal of 18 genotypes; value of 1975: dark column, value of 1976: light column

stable in 1976. Apparently the dry and hot climate of 1976 nullified the differences between mutants and between mutants and the initial line. Considering Figure 5 as a whole, we find that mutants with a high protein content (26-29%) in the first year, produced a similar high one in the second vegetation period. But in those cases where protein content ranged between 21 and 24% in 1975, the value is improved by about 5% in 1976.

The value 'protein content' as such is not of much importance in plant breeding. Here, the combination of yield (seeds per plant), kernel size and, in addition, the protein content, is of interest. This is why we calculated the protein production per plant for the genotypes analysed in the course of this work. Figure 6 represents these results; the values are related to the initial line.

The genotypes can be combined to two groups. Group 1: protein yield is the same in both the years or is slightly higher in the second one. None of these values reach those of the initial line in both the years. An improved protein percentage was found in mutant 419 (Fig. 5) but as the yield was strongly reduced in this mutant (Fig. 7), the protein yield is low. The protein percentage of the genotypes 176X, 123, 1201A, and 239CH is lower than that of the initial line in 1975 (Fig. 5), and though their yield improved (Fig. 7), the protein production was lower than in the initial line. In the second year the protein percentage of all these mutants improved considerably (Fig. 5). But combined to this either the yield or the kernel size is reduced. This results in a similar value for protein yield as in 1975 (Fig. 6).

Group 2 (Fig. 6, right part) shows clear differences in protein yield between the years for all the mutants. In the first 8 genotypes the protein yield is higher in the second year, while in the last one (mutant 489C) this value is reduced. In 1975, the protein production of all mutants, except of mutant 489C and perhaps mutant 251A, was less to that of the initial line due to a reduced yield (in mutants 12C, 486, 46C, 157A, 2314 – Fig. 7), a reduced kernel size (in mutants 12C, 105, 123) or a reduced protein content (in mutants 486, 2797). The reduced protein percentage of mutant 251A and 489C (Fig. 5) is compensated by an extraordinary high yield in this year (Fig. 7).

In the second year (1976), the protein production of the 8 mutants 12C-251A (Fig. 6, right part) is clearly better than in the preceeding year and with the mutants 486, 157A, 2797, 2314, and 251A the protein production is higher than with the initial line in the same year (Fig. 6). These high values are either due to the improved protein percentage or to an increased yield; in mutant 251A and 486 both these factors interact resulting in an extraordinary high protein yield. In contrast to these genotypes, mutant 489C shows a reverse behaviour: Protein yield is reduced in the second year (Fig. 6). Though the protein percentage is high in this vegetation period, a strongly reduced yield (Fig. 7) leads to the reduction in protein yield.

Preferently, mutants are compared to their initial line but in some cases it is of interest to compare them to each other. In Figure 1-4, mutants that differ from each other are joined by a line. With mutants of group 4 (Fig. 4) protein values are similar to each other, while in group 3 (Fig. 3) the protein value of mutant 486 is definitely different from the two other genotypes. As to genotypes in group 2, we see from Figure 2 that mutant 12C clearly can be distinguished from mutants 2797 and 251A concerning protein percentage. The mutants 157A, 239CH, and 1201A, which belong to group 1, are bifurcated and

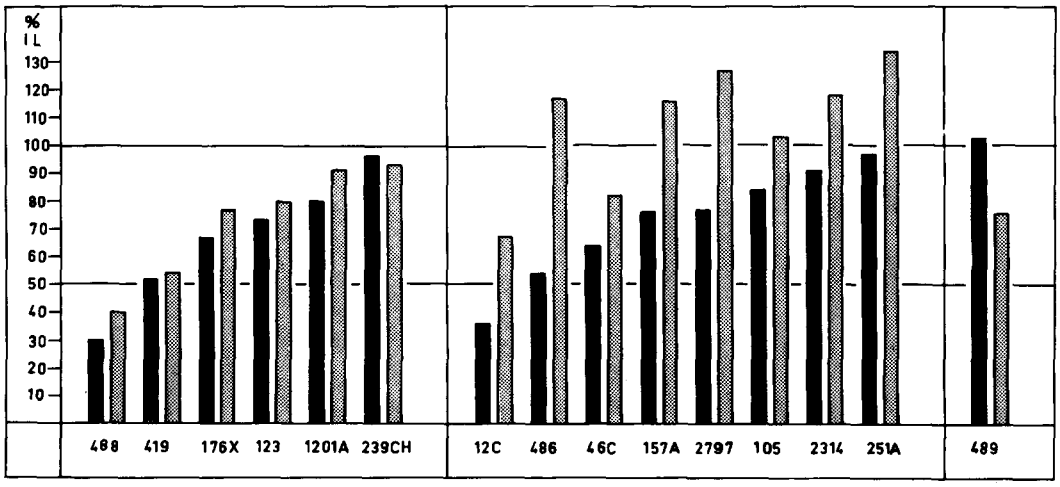

Fig. 6. Protein production per plant of 15 mutated genotypes, values related to those of the initial line. Value of 1975: dark column, value of 1976: light column

phenotypically very similar. From penetrance data it had been concluded that the respective genes are not identical (Gottschalk and Chen 1969, Gottschalk 1978). Our investigations point in the same direction. The protein value of 157A is significantly higher than both the values of 1201A and 239CH (Fig. 1). This means that mutant 157A is definitely different from 239CH and 1201A but we cannot decide whether 239CH and 1201A are identical or different in this connection.

The differences between the mutants are found only in the first year of investigation; in the second one no such differences exist between the mutants.

Discussion

Quantitative characters – such as seed protein content – show continuous distribution, the differences between mutant lines and between mutants and their initial line might be small and overimposed due to environmental interactions.

In former papers on seed protein content of our mutants, difficulties arose with the delimitation of mutants in connection with seed protein content. The most interesting mutant 489C, was a good example for these difficulties. In 1973, seed protein values were published (Mül-

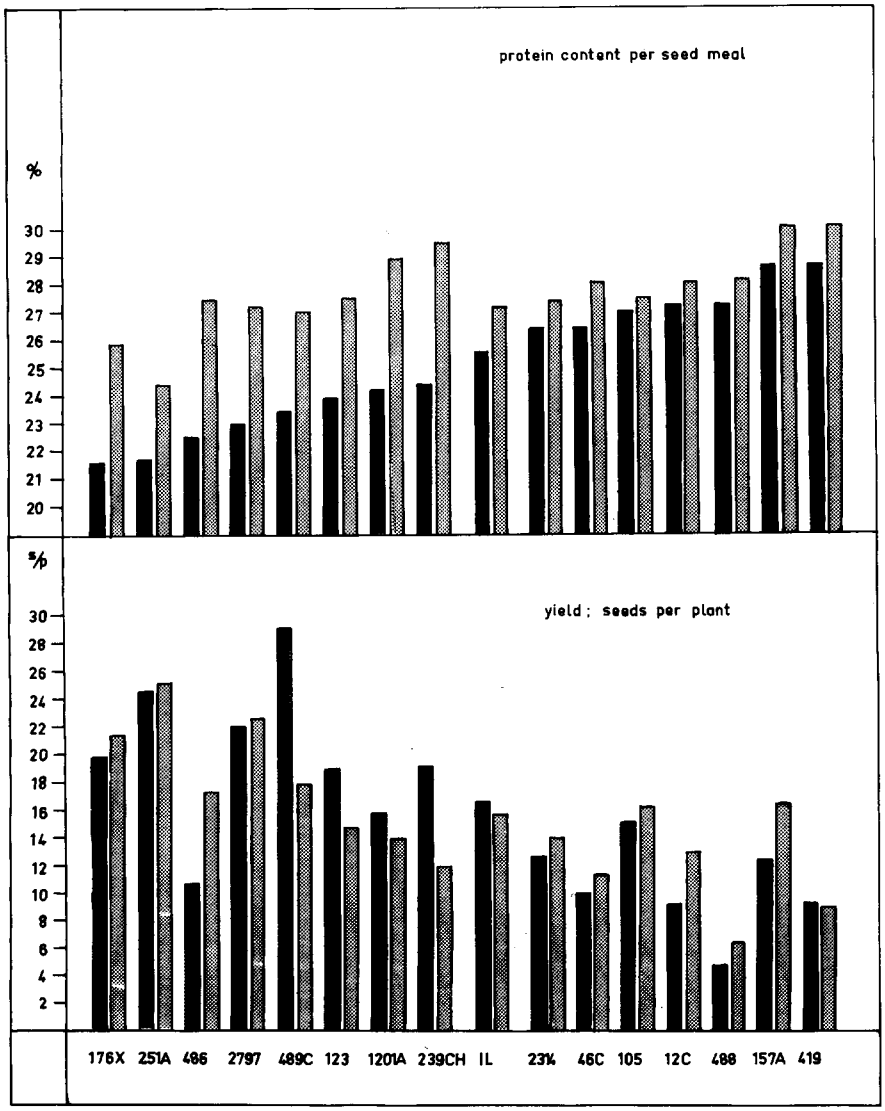

Fig. 7. Protein content per seed meal of 16 genotypes in 1975 (dark column) and 1976 (light column; upper part) and yield, seeds per plant, in the same years (lower part)

ler and Gottschalk 1973), which exceeded those of the initial line by about 20%, considering samples of 1969, 1970, and 1971. Examinations of this material in other years yielded different values: 88.4-109% (Gottschalk et al. 1975). In other words, we did not know whether this mutant was better or similar to the initial line or not. In all these cases, an average value of the corresponding genotype and year was the basis of this comparison, without taking into account the variation range of the protein percentage. Applying this method, differences between the genotypes, especially if they are small, are uncertain and cannot be verified. By using the biometric model described, we clearly can calculate the differences between the mutants. The presupposition for these calculations is the use of as many single values as possible, instead of averages. The preparation of these values – especially on the basis of protein estimation – is connected with a big consumption of time and work. In our case 25 values for each genotype per year were used. Under certain circumstances – as, for example, in our case with interactions of genotypic and year effects – a greater quantity of single values turned out to be necessary to get confidence intervals short enough to make a satisfying statement. This is why in plant and mutation breeding programs this method might not be practicable. In those cases the usual method using average samples and, as a result of this, average protein values, should be used as before. But after having selected some interesting mutants, the use of the described biometric computations might give an exact and reliable method to decide on the usefulness of a specific mutant. In this connection the amount of work and time seems worthwhile.

Considering the mutants as a whole, our results show that the genotypes react differently under various weather conditions. In 1975, the protein content of some of the mutants varies clearly from the corresponding values of the initial line. Environmental influences can be excluded as reason for these differences, therefore they must be genetically caused.

Apart from some genotypes we find a loose correlation between yield and protein content. A high yield is combined to a low protein value, and reverse. Similar findings were published earlier (Wolff 1975) and were found in other material, too (Kushihuchi et al. 1974; Scholz 1975; Narahari et al. 1976). In those cases the protein content directly might be dependent on yield. The mutated gene influences yield, and alterations of protein content are a secondary effect. It is supposed that protein, on the whole, is quantitatively changed. Unequal alterations and qualitative deviations of single fractions can be excluded as the protein-bound amino acid composition is completely unchanged in seeds of all genotypes mentioned in this paper (Wolff unpublished). Three genotypes escape this correlation: mutants 157A, 105 and 486. The first two mutants show a good yield, similar to that of the initial line, and combined to this, good protein values. With the third mutant a rather bad protein percentage is found, while the yield likewise is bad. This finding is so far of importance as many authors are convinced that the negative correlation between protein content and yield cannot be broken. Further investigations will be made into this problem.

Concerning protein content, the findings of 1976 are very surprising. None of the differences found in the preceding year occurred in this vegetation period. The uniformity of the protein values is astonishing. All values are increased and range about 27% protein per seed meal. This means that genotypes with a usually reduced protein content in seeds (23%) are improved considerably under extreme climatic conditions while mutants with a good protein situation are only slightly improved or not at all. It might have been possible that the increased protein content was due to a reduced yield in this extreme year, in a similar way as was discussed for values of 1975. But this we could not find. This might be another example that correlation between yield and protein percentage is not a strong rule. Mutants 489C, 239CH and 123 excluded, yield in 1976 increased or was similar to the values of 1975. Different reasons must be under discussion for this behaviour. The finding that yield in the second year (1976) was generally similar to that of 1975 was not expected. The dry and hot weather of 1976 was very unusual to our region and might have had a negative effect on a pea variety adapted to our temperate region. This behaviour might be explained with the origin of pisum. The Near East, including areas of Central Asia, the Mediterranian region and Abyssinia are supposed to be the home of peas. These regions, in general, are hot and dry. Apparently our pea variety is still adapted to such a climate: concerning yield, our peas produce the same quantity of seeds per plant under moderate as well as under extremely warm conditions. Apart from this general behaviour single mutants exist that seem to have lost this adaptability to both climates: Mutants 489C, 239CH and 123 showed better values in the cool and rainy year, while with mutants 12C, 486, and 157A better values are obtained in the dry and hot one. Apparently under influence of mutated genes, the adaptability of a genotype to various climatic conditions can be changed.

Concerning protein percentage, mutants showed uniformly good values under the hot and dry conditions of the second year. Genetically caused differences did not appear. One group of mutants (Fig. 5, left) showed a broad adaptability, and in 1975 as well as in 1976, good protein values are available. In the second group (Fig. 5, right), on the other hand, the cool climate reduced protein output, and only under the warm conditions are the values more or less good. In connection with this trait,

these mutants are not adapted to our temperate climate. The result is an improvement of protein yield in a considerable number of mutants in the second year.

Literature

Doll, H.; Koie, B.; Eggum, B.O.: Induced high lysine mutants in barley. Radiation Bot. **14**, 73-81 (1974)

Gebhardt, F.: Verteilung und Signifikanzschranken des 3. und 4. Stichprobenmoments bei normal verteilten Variablen. Biom. Z. **8**, 219-241 (1966)

Gottschalk, W.: Fasciated peas – Unusual mutants for breeding and research. J. Nuclear Agric. Biol. **6**, 27-33 (1977)

Gottschalk, W.: The dependence of the penetrance of mutant genes on environment and genotypic background. Genetica (in press) (1978)

Gottschalk, W.; Chen, R.: Die Penetranz mutierter Gene als begrenzender Faktor in der Mutationszüchtung. Z. Pflanzenzücht **62**, 293-304 (1969)

Gottschalk, W.; Müller, H.P.: Quantitative and qualitative investigations on the seed proteins of mutants and recombinants of *Pisum sativum*. Theor. Appl. Genet. **45**, 7-20 (1974)

Gottschalk, W.; Müller, H.P.: Monogenic alteration of seed protein content and protein pattern in X-ray induced *Pisum* mutants. In: Improving Plant Protein by Nuclear Techniques. Proc. IAEA, Vienna 201-212 1970

Gottschalk, W.; Müller, H.P.; Wolff, G.: Relations between protein production, protein quality and environmental factors in *Pisum sativum*. In: Breeding for Seed Protein Improvement using Nuclear Techniques. IAEA Vienna 105-123 1975

Gottschalk, W.; Müller, H.P.; Wolff, G.: Further investigations on the genetic control of seed protein production in *Pisum* mutants. In: Evaluation of seed protein alterations by mutation breeding, IAEA Vienna 157-177 1976

IBM-Corp PL/I-FORMAC Interpreter New York: Hawthorne (1967)

Kushihuchi, K.; Okada, M.; Watanabe, Sh.: Breeding for high protein rice by the use of artificially induced mutants. Gamma Field Symposia **13**, 37-45 (1974)

Lund, R.E.: Tables for an approximate test for outliers in linear models. Technometrics **17**, 473-477 (1975)

Mertz, E.T.; Bates, L.S.; Nelson, O.E.: Mutant gene that changes protein composition and increases the lysine content of maize endosperm. Science **145**, 279-286 (1964)

Müller, H.P.; Gottschalk, W.: Quantitative and qualitative situation of seed proteins in mutants and recombinants of *Pisum sativum*. In: Nuclear Techniques for Seed Protein Improvement, IAEA Vienna 235-253 1973

Narahari, P.; Bhatia, C.R.; Gopalakrishna, T.; Mitra, R.K.: Mutation induction of protein variability in wheat and rice. In: Evaluation of Seed Protein Alterations by Mutation Breeding, IAEA Vienna 119-127 1976

Proceedings of the IAEA: Approaches to Breeding for Improved Plant Protein. Vienna 1969. Improving Plant Protein by Nuclear Techniques Vienna 1970; Nuclear Techniques for Seed Protein, Vienna 1973; Breeding for Seed Protein Improvement, Vienna 1975; Evaluation of Seed Protein Alterations by Nuclear Breeding, Vienna 1976; Seed Protein Improvement by Nuclear Techniques, Vienna 1978

Putter, J.: Orthonormal bases of error spaces and their use for investigating the normality and variances of residuals. JASA **62**, 1022-1036 (1967)

Quednau, H.D.: Die Durchführung biometrischer Analysen mit Hilfe symbolverarbeitender Computersprachen. Habilitationsschrift Math.-Nat. Fak. Univers. Bonn 1978

Quednau, H.D.; Wolff, G.: Biometric evaluation of seed protein production of *Pisum* mutants and recombinants. In: Seed Protein Improvement by Nuclear Techniques 315-329. IAEA Vienna 1978

Scheffé, H.: A method for judging all contrasts in the analysis of variance. Biometrika **40**, 97-104 (1953)

Scholz, F.: Problems in breeding for high protein yield in barley. In: Barley Genetics III. Proc. 3rd Intern. Barley Genet. Symp. 548-555. Garching 1975

Singh, R.; Axtell, J.D.: High lysine mutant gene (hl) that improves protein quality and biological value of grain sorghum. Crop Sci. **13** 535-539 (1973)

Wolff, G.: Quantitative Untersuchungen über den Proteingehalt von Samen von *Pisum sativum*. Z. Pflanzenzücht. **75**, 43-54 (1975)

Wottawa, H.: Das 'allgemeine lineare Modell – Ein universelles Auswertungsverfahren. EDV in Medizin und Biologie **5**, 65-73 (1974)

Received June 5, 1978
Communicated by H. Stubbe

Dr. H.D. Quednau
Dr. G. Wolff
Institute of Genetics
University of Bonn
Kirschallee 1
53 Bonn (West Germany)