

TOWARDS BETTER CANOLA YIELD: A PRINCIPAL COMPONENTS ANALYSIS APPROACH

G.J. Lewis and A.T. Lisle

School of Land and Food, The University of Queensland, Lawes, Qld. 4345

Abstract

Principal component analysis was used to examine the relationships among phenological, yield component and growth traits of a set of canola lines developed by the inbred backcross technique. The lines were evaluated in the field at East Beverley, Western Australia. This showed that 78 % of the variation in the traits was explained by the first four vectors, which were concerned with growth/ plant size, post-anthesis growth, phenology and yield component compensation, respectively. Seed yield was used as the independent variable in a multiple regression. Eighty percent of the variation in yield was explained by the growth/plant size vector and a further 9 % was explained by the yield component compensation vector. The growth/plant size vector also effectively explained the yield of a subset of nine of the original lines grown at the same site two years later

Key words: Canola, principal component analysis, yield.

Principal component analysis has been shown to be a useful technique in clarifying the relationships among traits in complex multi-trait systems such as the growth of a crop. The purposes of this paper are:

- to use principal component analysis to clarify the relationships between the physiological and phenological traits of a set of canola lines;
- to describe the relationship of the vectors extracted with seed yield in a medium rainfall, Mediterranean environment; and,
- to show how principal component analysis can be used to assist in identification of selection criteria and selection of lines within a breeding programme.

Materials and methods

A set of inbred backcross lines (B3S3) were developed by crossing *Brassica napus* cv. Wesbrook with the early flowering Victorian line RU2. Forty eight of these lines selected at random along with Wesbrook and Eureka were tested in a randomised complete block experiment at East Beverley, Western Australia which was sown on June 8, 1989. Plots were 5 m long and spaced 0.18 m apart. A second experiment containing Wesbrook and 8 of the inbred backcross lines from the initial experiment was planted in 25 m x 1.42 m plots on June 14, 1991.

Characters measured were stem elongation (days; SE), stem elongation to first open flower (days; SE-FOF), first open flower - physiological maturity (days: FOF-M), dry matter at first open flower (g/m^2). DMF), dry matter increment between first open flower and maturity (g/m^2 : DMI), crop growth rate between sowing and first open flower ($\text{g/m}^2/\text{d}$: CGR_F), crop growth rate between first open flower and physiological maturity ($\text{g/m}^2/\text{d}$: CGR_{PF}), seed yield (g/m^2), harvest index (%: HI), distribution index (%: DI), number of pods per square metre, number of seeds per pod and thousand seed weight (TSW).

The data set was analysed using principal component analysis (4). Multiple regression analysis was used to relate the values of each of the principal components extracted for each line to seed yield.

Results

Variability within the data

Table 1: Percentage of total variance within the data accounted for by the first four vectors derived from the principal component analysis (eigenvalues) and coefficient of determination (r^2) with seed yield.

Vector	Proportionate Contribution to Variance ^a	r^2
1	28.3	0.81
2	20.2	0.04
3	15.8	0.02
4	13.4	0.08

^a The first four vectors accounted for 77.8 % of the variance within the data set

Four principal components were identified which accounted for more than 10 % of the variation within the data set (Table 1). Cumulatively, these four vectors accounted for over 75% of the variation.

Interpretation of the principal components

Table 2: Standardised loadings on the first four principal components

Parameter	Principal Components			
	1	2	3	4
SE	-0.180	-0.280	0.584	0.066
SE-FOF	0.188	-0.177	-0.559	-0.130
FOF-M	0.054	0.551	-0.226	0.041
DMF	0.420	0.232	0.214	-0.212
DMI	0.421	-0.282	-0.016	0.295
CGR _F	0.409	0.294	0.178	-0.198
CGR _{PF}	0.387	-0.364	0.032	0.277
HI	0.338	0.106	0.093	0.235
DI	0.116	0.326	0.247	-0.065
Pods/ m ²	0.321	-0.246	0.145	-0.436
Seeds/ pod	0.057	0.223	0.050	0.692
TSW	0.156	-0.081	-0.352	0.008

The first principal component extracted (Prin. 1) can be regarded as a growth or size vector (Table 2). It has high loadings for DMF, DMI, CGR_F, CGR_{PF} and harvest index. There is some indication from this vector that harvest index is higher when DMF and DMI are larger ($r = 0.33^*$, 0.41^{**}). The second principal component extracted (Prin. 2) is a reproductive growth vector. It has high loadings for FOF-M, CGR_{PF} and DI. In lines with a high crop growth rate after first open flower the duration of the period between first open flower and physiological maturity is shorter ($r = -0.38^{**}$). The third principal component (Prin. 3) is a developmental vector (Table 2) which shows that early onset of stem elongation is associated with long duration of the period between stem elongation and first open flower ($r = -0.67^{***}$). The duration of the period between stem elongation and first open flower is also positively related with thousand seed weight ($r = 0.32^*$). Finally the fourth principal component extracted (Prin. 4) is a yield component compensation vector (Table 2). When the number of pods per square metre is high number of seeds per pod will be low and vice versa ($r = -0.53^{***}$).

These four principal components can be used as new traits containing information from all of the 12 traits originally measured or alternatively the loadings on each trait can be used as the basis for constructing

new indices which have a simpler biological interpretation and which require measurement of fewer initial traits. Prin. 1 could be represented by total dry matter, a trait combining DMF, DMI, CGR_F and CGR_{PF} with the loss of relatively little information. Likewise Prin. 2, Prin. 3 and Prin. 4 can be represented by the ratio of DMI to DMF, SE and the ratio of pod number to seeds per pod respectively.

Impact on seed yield

Prin. 1 accounts for over 80% of the variation in seed yield among the lines and Prin 4. accounts for a further 8% (Table 3). However, these two vectors do not explain as much of the variation in seed yield among the lines as total dry matter, the trait representing Prin. 1., alone (90% C/F. 89%).

Stability of the relationship across environments

Yield predictions of a subset of nine lines from the original experiment, grown in 25 m x 1.4 m plots at the same site, two years after the original experiment, using values of Prin. 1 calculated from the loadings in the original experiment were highly correlated ($r = 0.94^{**}$) with the actual yields and the slope and y-intercept were not significantly different from 1 and 0, respectively. Therefore, it can be concluded that the predictive value of the most important principal component is stable across environments.

Discussion

Principal component analysis in combination with agglomerative cluster analysis is the recommended technique for studying the interactions between geno-types, environments and attributes (3). Generally the emphasis of these studies is on defining the structures within the genotype, environment and attribute data and their interrelationships, rather than interpretation of the components themselves (1). The current study emphasises interpretation of the vectors in order to gain an understanding of the processes contributing to differences in yield among the lines and to assist in identification of selection criteria and lines which are worthy of further evaluation.

The results show that high dry matter production is the primary factor determining high canola yield in a medium rainfall mediteranean type environment. This contrasts with the situation with cereals where improvement in harvest index has been responsible for the majority of improvement in seed yield (5). Selection for high dry matter production either directly or using the growth/plant size vector identified (Prin. 1) will be effective in improving seed yield because there is a positive relationship between total dry matter and harvest index. However, it will be simpler and more cost effective to measure total dry matter alone, rather than the other traits which are necessary to obtain values for Prin. 1.

Lines with high values of total dry matter (or Prin. 1) should be retained for further testing in this environment. This strategy is vindicated because one line WBRU-9, which was of only average yield in the initial experiment, but had high value of Prin. 1 was significantly higher yielding than any of the other lines in the subsequent large plot experiment (2).

If further crosses were to be conducted between lines from within this population it would be best to choose lines that had high total dry matter (Prin. 1), but contrasting values of Prin. 4, which was concerned with distribution of yield components. This would maximise the possibility of identifying new lines combining high total biomass, high pod number and a large number of seeds per pod.

Another option is to use principal components analysis on several populations grown in the same environment to determine if the association between seed yields and the vectors extracted was the same or different for the different populations. If there is a different relationship with seed yield for the different populations opportunity exists to develop new lines containing the characteristics associated with high yield in each population by using the best line from each cross as parents in a new crossing programme

Conclusions

Principal components analysis was an effective technique for gaining an understanding of the relationships between traits in a complex multi-trait set of data derived from a breeding population. However, in this case selection of lines on a single trait (total dry matter) would have been equally effective. Use of values of Prin. 1 and Prin. 4 would have assisted the breeder in choosing parents for subsequent crosses.

Acknowledgments

The data on which this paper is based was obtained from the senior authors Ph.D. project at the University of Western Australia, which was funded by the Oilseeds Research Council of Australia. The senior author would also like to acknowledge the assistance and support of his supervisor, the late Prof. Noel Thurling during this project.

References

1. Basford, K.E, Kroonenberg, P.M. and Cooper, M 1996. In "Plant Adaptation and Crop Improvement". Edited by M. Cooper and G. Hammer. CAB International, Wallingford, U.K. pp. 291 -306.
2. Lewis, G.J. 1997. Proc. Eleventh Aust. Res. Ass. Brassicas. Edited by G. Walton. Perth, pp. 74 - 78.
3. McLaren, C.G. 1996. In "Plant Adaptation and Crop Improvement". Edited by M. Cooper and G. Hammer. CAB International, Wallingford, U.K.. pp. 225 - 242
4. Morrison, D.A. 1967. "Multivariate Statistical Methods". McGraw-Hill, New York, pp. 415.
5. Siddique,K.H.M, Belford, R.K , Perry, M.W. and. Tennant, D.T. 1989. Aust. J. Ag. Res. 40, 473 - 487.