RESEARCH ARTICLE

Marker-aided genetic divergence analysis in *Brassica*

V. ARUNACHALAM*, SHEFALI VERMA, V. SUJATA and K.V. PRABHU

Division of Genetics, Indian Agricultural Research Institute, New Delhi 110 012, India

Abstract

Genetic divergence was evaluated in 31 breeding lines from four *Brassica* species using Mahalanobis' D^2. A new method of grouping using D^2 values was used to group the 31 lines, based on diagnostic morphological traits (called morphoqts). Isozyme variation of the individual enzymes esterase and glutamate oxaloacetate was quantified by five parameters (called isoqts) developed earlier. Grouping by the same method was also done based on the isoqts, and the grouping by isozymes was compared with that by morphoqts. Overall, there was an agreement of 73% suggesting that isoqts can be used in the choice of parents and also first stage selection of segregants in the laboratory. It was suggested that such an exercise would help to take care of season-bound and field-related problems of breeding. The new isozyme QTs, within lane variance of relative mobility and relative absorption, accounted for about 50% of the total divergence. The utility of the new method and isoqts in cost-effective breeding were highlighted.

[Arunachalam V., Verma S., Sujata V. and Prabhu K. V. 2005 Marker-aided genetic divergence analysis in *Brassica J. Genet.* **84**, 123–130]

Introduction

Brassicas are an important group of edible oil crops in India, with four widely used species being *B. juncea*, *B. napus*, *B. carinata* and *B. campestris*. While *B. juncea* and *B. campestris* are the most popular edible oil crop species of the genus, *B. napus* has also been gaining importance as productive lines have been developed in the recent past, and *B. carinata* is often used as a donor parent in hybridisation programs. However, all the four species are winter crops and grow optimally only in cool temperatures.

Classical breeding for quantitative trait (QT) improvement uses a grouping based on genetic divergence for selecting parents for initiating hybridization. In more recent times, assessing genetic divergence using a variety of molecular markers has become common. However, along with DNA markers, isozyme markers have also been found very effective in genetic divergence studies (e.g. Zvingila *et al*. 1998). Plant breeders, however, typically use diagnostic morphometric traits to measure genetic divergence. A popular measure of genetic divergence is the multivariate distance statistic, D^2 due to Mahalanobis (for detailed account, see Rao 1952; Arunachalam 1981). However, this measure of genetic divergence requires morphometric data from crops usually grown in the field. In such cases, biotic and abiotic stresses could interfere with the performance of the crop. Moreover, crops could be seasonal and restricted to a particular season in a year, implying that repeat data could need a long waiting time. Where heterozygous plants are involved, plants selected for hybridisation will require regeneration from saved seeds, as otherwise the progeny of selected heterozygous plants would segregate in the subsequent generation. This would elongate the process by one more season and become time-, cost- and labour-intensive.

The covariation of isozymes with environment makes it possible to study their variation in a manner similar to that used in studies of morphological variation. We have earlier characterized isozyme variation as a set of five isozyme QTs (hereafter abbreviated as isoqts)—number of bands, relative mobility (RM), relative absorption (RA), and the within lane variance of RM and RA, respectively (WLM and WLA) (Arunachalam *et al*. 1996). Genetic

*For correspondence. E-mail: vaiarunachalam@hotmail.com, varunachalam@mssrf.res.in

Keywords. Genetic divergence; *Brassica*; isozymes; Mahalanobis' D^2; Grouping.

divergence based on isozyme markers could then be measured using D^2 based on the five isoqts. A new grouping method was devised using all possible D^2s. To compare the new methodology with the popular dendrogram grouping usually followed in marker studies, dendrograms were also constructed using the multivariate divergence matrix for morphoqts and isoqts. In this paper, we sought answers to the following questions:

(i) Time-tested breeding is based on phenotypes that are grouped using divergence based on diagnostic morphoqts. What is the magnitude of agreement of grouping of the same test lines by one or more isoqts?
(ii) Is the magnitude of agreement sustainable over repeats of experiments across different environments like different growing seasons, and different genetic status of material (open-pollinated or selfed)?
(iii) Can there be a preferential identification of the test lines by isoqts?
(iv) How does the new methodology compare with the popular dendrogram grouping?
(v) How would isoqt based grouping help in plant breeding?

Material and methods

Marker-aided divergence studies used 31 untested lines developed from a large germplasm (including a few released high yielding lines) from the four *Brassica* species referred to earlier (table 1). The material was grown in the field with optimum cultural practices in a randomised block design at the Indian Agricultural Research Institute, New Delhi, in four experiments designated as experiment A1 to A4, over three seasons. Experiments A1 and A4 had two, and experiments A2 and A3 three replications. In experiments A1 and A2, the material was open-pollinated. Seeds selfed once were used to raise the material in experiments A3, whereas seeds selfed twice in succession were used for experiment A4.

Observations on flowering time, total number of branches, seed yield per primary branch, seed yield per plant and harvest index, based on means of samples of three to five plants were used as diagnostic morphological traits (to be abbreviated as morphoqts), except for experiment A1, in which plant height instead of harvest index was measured. The enzymes esterase (EST) and glutamate oxaloacetate (GOT) were used for isozyme analysis following the procedure of vertical gel electrophoresis described in Dadlani and Varier (1993). For isozyme assays, plants were sampled in the field and two plants per accession per replication were tagged.

Gel patterns were scanned and the relative mobility and relative absorption were estimated by a laser densitometer (Pharmacia LKB, Ultrascan XL). A gelscan XL software supplied by the company helped to record the data in a DIF format. A software was developed to retrieve from the DIF format, the data on number of bands, RM and RA and to compute WLM and WLA to get the five isoqts for divergence analysis.

A new methodology, developed and tested using a number of crops, was used in this study. In this method, the mean (m) and standard deviation (s) of all the D^2 values were computed. Keeping in mind the property of a normal distribution that 68% of the total area lies between (m–s) and ($m + s$), the D^2 values were assigned to one of four divergence classes (DC) as shown in figure 1. The advantages of grouping the D^2 values uniformly into four divergence classes are:

- The norms for grouping are self-set and therefore result in unique groups.
- It facilitates comparison of experiments repeated over locations, seasons etc. on the four divergence classes uniformly set in each case. When necessary, 6, 8, 10 etc. groups could also be set using the same logic. Our experience suggests that four groups are adequate.
- It enables appropriate choice of parents, a major step in breeding.
- Where there is scope, experiments across crops and global environments, can also be evaluated for convergence of concepts e.g. parental divergence *vis-à-vis* F_1 heterosis.
- The concept of assigning D^2s to divergence classes made it possible to allot relative scores to every one of the 31 lines; for example, line 1 could occur in various D^2 combinations in DC1 to DC4 that can be enumerated. In fact there would be a 31×4 matrix of values (say X_{i1}, X_{i2}, X_{i3}, X_{i4}; $i = 1...31$) corresponding to the 31 lines and 4 divergence classes.

Theoretical and practical studies over the past two decades have revealed that it is prudent to select parents from the divergence classes DC3 and DC2 to enhance the chances of F_1 heterosis (Arunachalam 1977; Arunachalam and Bandhopadyay 1984; Arunachalam *et al*. 1982, 1984). This result provided the logic to assign further weights. Since DC3 and DC2 would have intermediate divergence values preferred for heterosis, a weight 1 was given to them. DC1 would have extremely high values and be unlikely to give heterosis; hence, a weight of –1 was given to DC1. DC4 having low D^2 values may occasionally give heterosis and, hence, received a weight of 0. When X_{i1}, X_{i2}, X_{i3}, X_{i4}; ($i = 1...31$) for the four DCs are multiplied by their weights, we get an alignment score for each entity. For instance, let the frequency of occurrence of line 1 in D^2 combinations be 5 in DC1, 3 in DC2, 4 in DC3 and 3 in DC4. The alignment score of line 1 would then be given by $[(5 \times -1) + (3 \times 1) + (4 \times 1) + (3 \times 0)] = 2$ (see Arunachalam *et al*. 1998 for further details). The alignment scores thus obtained were set in four final groups, Group 1 to Group 4 (table 2) using the same logic as that for setting four divergence classes. We

note that logically Group 1 should represent the best and Group 4 the least in relative merit.

Results

For isoqts to be evaluated further for first stage selection, the contribution of isoqts of EST and GOT to genetic divergence must be comparable with that of morphoqts. It was found that flowering time and single plant yield among morphological traits contributed ≅ 50% to divergence. Similar contributions were made by within lane variance of relative mobility (WLM) and of relative absorption (WLA) together (table 3). The traits flowering time, single plant yield and total number of branches accounted for ≅ 68%. In comparison, relative mobility, relative absorption and their within lane variances contributed over 85% to divergence, both for EST and GOT. Number of bands contributed the least in almost all the experiments both for EST and GOT (table 3).

Table 1. Lines identified by isoqts using enzymes esterase (E) and glutamate oxaloacetate (G) in the various experiments A1...A4.

Species/lines	V	A1		A2		A3		A4
B. juncea								
5226–5–1S	1	E	E	–	E	G	–	G
5156–3–1S	2	E	E	G	E	G	E	G
5063–1–1S	3	E	E	G	E	G	E	G
5044–1–1S	4	E	E	G	E	G	E	G
5123–3–1S	5	E	–	G	–	–	E	G
Pusa Bold	6	E	–	G	E	–	–	G
Krishna	7	E	E	G	E	G	E	G
Kranti	8	E	–	–	–	G	E	G
Varuna	9	E	E	G	E	G	–	–
5159–6–1S	10	E	E	G	E	G	E	G
5010–1–1S	11	–	–	G	E	G	E	–
5057–3–1S	12	E	E	G	E	G	E	G
5067–1–1S	13	E	E	G	–	–	E	G
5115–1–1S	14	E	E	G	E	G	–	–
5122–13–1S	23	E	E	G	E	G	–	G
5060–11–1S	24	E	E	–	–	G	E	G
5224–11–1S	25	E	–	G	–	G	E	G
5127–11–1S	26	–	–	G	E	G	E	–
Total	18	16	12	15	13	15	13	14
B. napus								
3012–21–1S	22	–	E	G	E	G	E	–
3023–15–1S	27	–	–	–	E	G	–	G
3019–17–1S	28	E	–	–	E	G	–	–
Total	3	1	1	1	3	3	1	1
B. carinata								
2019–2–1S	15	–	–	–	–	–	–	–
2062–1–1S	16	–	–	–	–	–	–	–
2012–5–1S	17	E	–	–	–	G	–	–
2014–2–1S	18	–	–	–	–	–	–	–
2019–11–1S	29	–	–	–	–	–	E	G
Total	5	1	0	0	0	1	1	1
B. campestris								
1073–2–1S	19	–	E	G	E	–	–	G
1042–1–1S	20	E	E	–	–	–	E	–
1076–2–1S	21	E	–	G	E	–	–	G
1038–12–1S	30	E	–	G	–	–	–	–
1009–11–1S	31	E	E	G	E	–	E	–
Total	5	4	3	4	3	0	2	2
Overall Total	31	22	16	20	19	19	17	18

V, Serial number of lines.

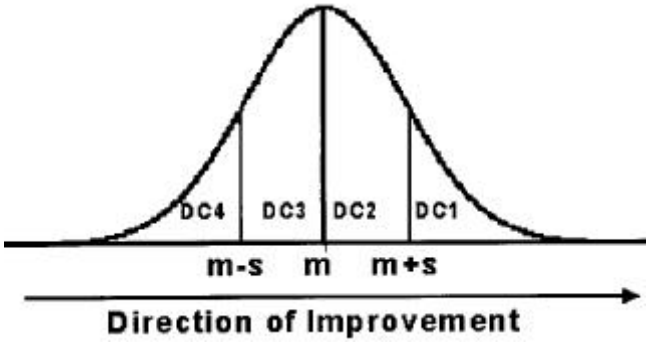


Figure 1. Divergence classes set by mean (*m*) and standard deviation(*s*) of divergence (D^2) values.

Table 2. Concordance between grouping of alignment scores based on morphoqts and esterase isoqts in *Brassica* – Expt. A1.

Group	BSE	EST	Common
1	0	0	0
2	24	24	20
3	3	3	1
4	4	4	1
Total	31	31	22 = 71%

BSE, Base morphoqts; EST, Esterase isoqts.
Entries indicate the frequency of occurrence.

Multivariate grouping by morphoqts and isoqts did not show any mix-up of lines of one species with any other (data not shown), thereby confirming the robustness of the methodology. Two *B. juncea* lines, 5057 and 5122 belonged to the Genetic Group 2 in our analysis. They produced F_1 heterosis when crossed with a released variety, Pusa Jaikissan and another *B. juncea* line, 5756 (data not shown). This observation suggests that parents selected from top ranking genetic groups would give heterosis, though more evidence is needed for generalization.

Interestingly EST identified a high per cent of lines ($\cong$ 90) of *B. juncea* in all the experiments followed by GOT in general. Isoqts (and that too EST only) identified less often the released but closely related varieties, Pusa Bold and Varuna out of the four tested in Expts. A1 to A4. In comparison, they identified the varieties, Krishna and Kranti more often (table 1). The number of lines of the other species used in these experiments was low; yet isoqts could identify several of them (table 4). However, EST OR GOT OR BOTH symbolised by (EST $\cup$ GOT) identified 73% of the 31 lines used in this study (table 4), making their performance as a diagnostic measure superior to that of either GOT or EST alone. Since, both the enzymes can easily be assayed, it would be simple to collate lines identified by (EST $\cup$ GOT). On a relative scale, EST and GOT individually identified more than 70% of lines in *B. juncea* followed by *B. campestris*, *B. napus* and *B. carinata*. The utilisation of those species in India

Table 3. Per cent contribution of morphoqts and isoqts to genetic divergence in *Brassica* in experiments A1...A4.

QT		A1	A2	A3	A4	Mean
Morphological	MFT	24.6	36.6	32.2	33.0	39.6
	NTB	15.4	17.8	11.7	15.1	14.1
	NPB	6.4	10.0	8.7	7.6	8.5
	YPB	12.7	12.0	11.7	15.2	13.3
	SPY	15.7	18.4	12.0	14.9	15.0
	HI	–	5.2	23.7	14.2	12.7
	HT	25.2	–	–	–	–
EST	BD	19.1	15.5	12.6	15.3	11.7
	RM	23.1	22.1	24.8	26.0	20.8
	RA	13.8	9.9	23.6	10.3	19.0
	WLM	29.2	30.9	18.2	24.1	27.9
	WLA	19.8	21.6	20.8	24.3	20.6
GOT	BD	–	9.7	8.9	16.0	15.5
	RM	–	28.9	24.3	16.4	23.0
	RA	–	16.7	12.8	18.8	15.2
	WLM	–	22.8	29.0	25.2	25.1
	WLA	–	21.9	25.0	23.6	21.2

MFT, mean flowering time; NTB, number of total branches; NPB, number of primary branches; YPB, average seed yield per primary branch; SPY, single plant seed yield (g); HT, plant height (Expt. A1 only); HI, harvest index (Expts. A2..A4); BD, number of bands; RM, relative mobility; RA, relative absorption; WLA, within lane deviation of RA; WLM, within lane deviation of RM.

is also in the same order. However, the degree of identification of *B. carinata* lines by EST and GOT was low while that of *B. napus* was a little better (table 4).

In general, the trend of results was similar across the four individual experiments A1 to A4, despite fluctuations in the per cent of lines identified. Overall per cent identification of the 31 lines was similar across experiments. Noting that experiments A3 and A4 used inbred material, we can indirectly infer that the degree of homogeneity of the material was high in all the experiments. A simple ranking of the efficiency of enzymes in identifying the pattern of genetic divergence (table 5) placed (EST ∪ GOT) at the top followed by GOT and then EST. The difference in ranking between GOT and EST (7 and 8) was, however, quite marginal.

In comparison, the classical approach of dendrogram grouping gave 7 groups under morphoqts (figure 2) and 10 groups under esterase isoqts (figure 3). Though the identification of esterase isoqt groups that correspond with morphoqt groups is not clear-cut, we find at most only seven corresponding groups and have to leave out three groups from esterase isoqts. However, Line 17 from *B. carinata* and Line 26 from *B. juncea* forming single entry groups (figures 2, 3) did correspond without doubt implying that the status of the two lines could be clearly identified by esterase isoqts. In contrast to the dendrogram approach, the methodology used in this paper arranged the 31 lines into 4 groups in descending order of merit based on morphoqts and isoqts. Therefore, morphoqt groups 1 to 4 can be compared with esterase isoqt groups 1 to 4 (table 2), yielding a 71% tally for esterase isoqt grouping with morphoqt grouping (table 1).

Discussion

The present study not only showed high agreement between morphoqt and isoqt grouping but also confirmed that (EST ∪ GOT) could identify 73% of the 31 lines consisting of four species of *Brassica*. Moreover, this result was found to hold over four repeats of experiments across three seasons, and both open-pollinated and selfed material. Thus, the results suggest that one can assay *Brassica* lines (and possibly germplasm too) using isoqts of EST and GOT with high confidence.

What are then specific advantages of a divergence analysis using identified isozymes, as compared to traditionally used morphoqts?

- Isozyme assay can be done in the seedling stage itself (over a period of four weeks as we found in our earlier unpublished studies, data not reported, providing enough time for a large number of assays over samples and replications) without affecting the growth of plants.
- Plants (including heterozygous ones) can be selected without losing a season and crosses with heterozygous plants be made in the same growing season (as selection can be made on isoqts much before flowering).
- It is possible to identify F_1 heterosis, using isoqts from enzyme assay using seedlings raised in plots of a screen house. Then, a large F_2 population of such

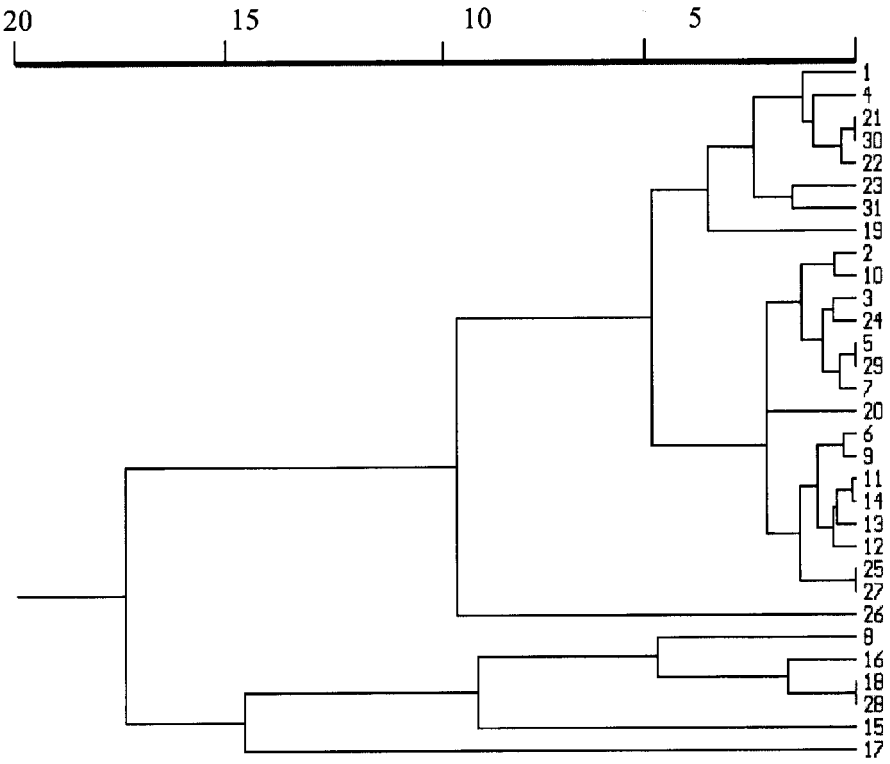
Table 4. Lines identified by EST, GOT or their combination in *Brassica*.

		A_1			A_2		A_3		A_4		Over A_1 to A_4	
Species	Enzyme	nt	nz	p	nz	p	nz	p	nz	p	Nz	p
B. juncea	EST		16	89	12	67	13	72	13	72	54	75
	GOT	18			15	83	15	83	14	78	44	81
	EST∪GOT		16	89	17	94	16	89	16	89	65	90
B. napus	EST		1	33	1	33	3	100	1	33	6	50
	GOT	3			1	33	3	100	1	33	5	56
	EST∪GOT		1	33	1	33	3	100	2	67	7	58
B. carinata	EST		1	20	0	0	0	0	1	20	2	10
	GOT	5			0	0	1	20	1	20	2	13
	EST∪GOT		1	20	0	0	1	20	1	20	3	15
B. campestris	EST		4	80	3	60	3	60	2	40	12	60
	GOT	5			4	40	0	0	2	40	6	40
	EST∪GOT		4	80	5	100	3	60	4	80	16	80
Overall	EST		22	71	16	52	19	61	17	55	74	60
	GOT	31			20	65	19	61	18	58	57	61
	EST∪GOT		22	71	23	74	23	74	23	74	91	73

nt, no. of lines tested; nz, no. identified; p, percentage; EST∪GOT, Either EST or GOT or both.

Table 5. Simple ranking of enzymes on their marking efficiency of divergence in *Brassica*.

	A2	A3	A4	Total	Final rank
EST	3	2	3	8	IIII
GOT	2	3	2	7	II
EST∪GOT	1	1	1	3	I



Group	Lines
1	1 4 21 30 22 23 31 19
2	2 10 3 24 5 29 7
3	20
4	6 9 11 14 13 12 25 27
5	26
6	8 16 18 28 15
7	17

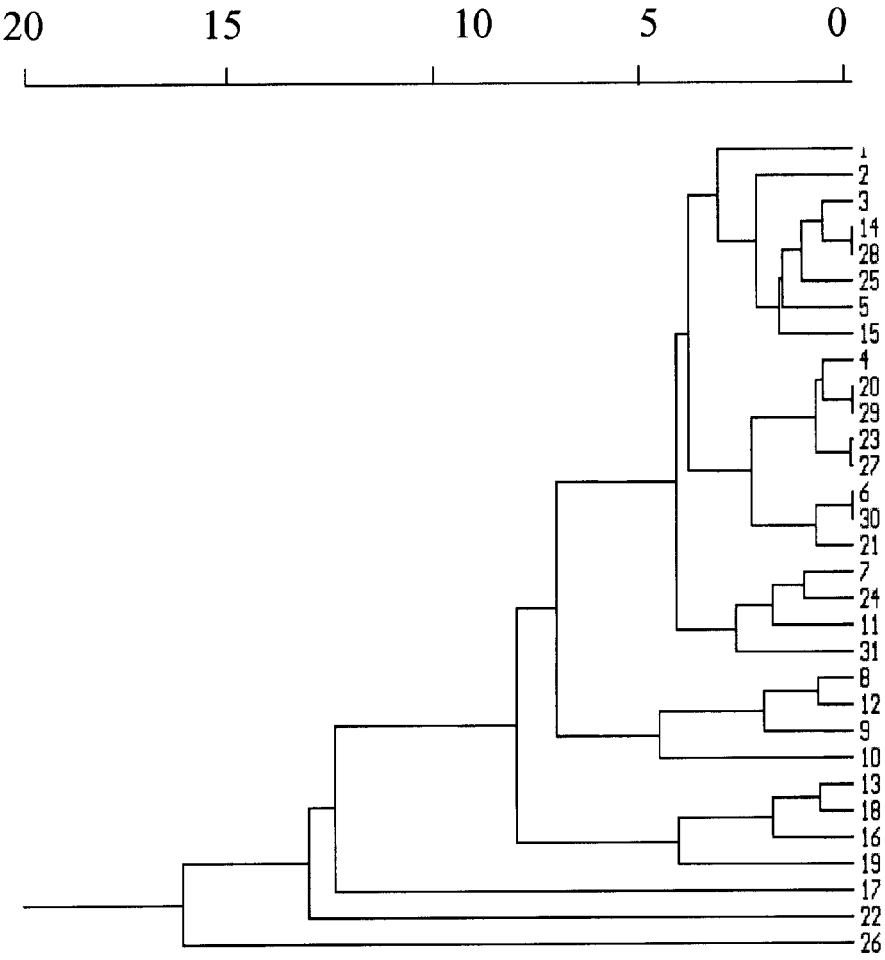
Figure 2. Dendrogram based on six morpholoqts—Expt. A1.

crosses only may be grown selectively, and a large number of F_2 plants assayed using select isoqts and prospective productive recombinants identified in advance. Only those plants can be grown to maturity in the field allowing great saving in growing space, invested time, cost and possibility to score all plants for diagnostic morphoqts in the field.

- A major advantage would be in selecting for productivity without interference by biotic and abiotic stresses in the field. Selected productive recombinants can be cycled through selection for stresses later.

With our present knowledge, we could not find a way to use multivariate D^2 statistic to measure divergence using molecular marker data. If this were possible, it would reinforce isozyme data to provide a better picture of genetic divergence. It would then be possible to compare grouping based on combined isozyme and molecular

marker data with morphological grouping. Nevertheless, the new methodology used to assess the material of experiments A1 to A4 appears to have specific advantages in parental choice, recombinant selection etc. While the present study used only two isozymes, the results were based on 1240 individual isozyme gel assays (taking into account, the no. of expts, the no. of replicates, the no. of plants sampled with each isozyme assay being mean of two individual readings) and suggested that they had high association with morphological traits. In case of molecular markers also, we need to use only those that have association with a diagnostic set of morphological traits defining phenotypic performance in field. Ideally, the efficiency of marker grouping and its utility needs confirmation in many crops, locations and seasons. To make this feasible, molecular biologists, quantitative geneticists and practical plant breeders should work as a team for evolving new paradigms of plant breeding.



Group	Lines
1	1
2	2 3 14 28 25 5 15
3	4 20 29 23 27
4	6 30 21
5	7 24 11 31
6	8 12 9 10
7	13 18 16 19
8	17
9	22
10	26

Figure 3. Dendrogram based on five esterase isoqts—Expt. A1.

References

Arunachalam V. 1977 Heterosis for characters governed by two genes. *J. Genet*. **63**, 15–24.

Arunachalam V. 1981 Genetic distance in plant breeding. *Indian J. Genet*. **41**, 226–236.

Arunachalam V., Bandyopadhyay A., Nigam S. N. and Gibbons R. W. 1982 Heterotic potential of single crosses in groundnut (*Arachis hypogaea* L.). *Oleagineux* **37**, 415-420.

Arunachalam V. and Bandyopadhyay A. 1984 Limits to genetic divergence for occurrence of heterosis – Experimental evidence from crop plants. *Indian J. Genet*. **44**, 548–554.

Arunachalam V., Bandyopadhyay A., Nigam S. N. and Gibbons R. W. 1984 Heterosis in relation to genetic divergence and specific combining ability in groundnut (*Arachis hypogaea* L.). *Euphytica* **33**, 33–39.

Arunachalam V., Prabhu K. V. and Sujata V. 1996 Isozyme parameters in genetic evaluation in *Brassica* – A tribute to Haldane's vision on the synergy between biochemistry and genetics. *Curr. Sci*. **70**, 1062–1065.

Arunachalam V., Prabhu K. V. and Sujata V. 1998 Multivariate methods of quantitative evaluation for crop improvement: conventional and molecular approaches revisited. In *Crop productivity and sustainability—shaping the future*. *Proc. 2nd International Crop Science Congress* (eds. V. L. Chopra, R. B. Singh and A. Varma), pp. 793-807. Oxford & IBH Publishing Co., New Delhi.

Dadlani M. and Varier A. 1993 Electrophoresis for variety identification. *Tech. Bull*. Division of Seed Science and Technology, Indian Agricultural Research Institute, New Delhi.

Rao C. R. 1952 *Advanced statistical methods in biometrical research*. John Wiley & Sons, New York.

Zvingila D., Kleizaite V. and Popendikyte V. 1998 Analysis of RAPD and isozyme (SOD) polymorphism in cereals. *Biotechnologija augalu selekcijoje, spalio*, **8–9d**, 33–37 (Lithuanian with English Abstract).

Received 18 August 2004; in revised form 6 May 2005