

Computer simulation of family selection schemes suitable for kale (*Brassica oleracea* L.), involving half-sib, full-sib and selfed families

J. E. Bradshaw

Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland

Received January 24, 1984; Accepted June 25, 1984

Communicated by Å. Gustafsson

Summary. Three recurrent selection schemes suitable for kale (*Brassica oleracea* L.), involving half-sib (HS), full-sib (FS) and selfed (S) families, were compared by computer simulation. All combinations of 6, 12 and 24 families selected, out of 120 and 240 assessed, were investigated for a range of genetical models. Selection was simulated for 20 generations from an initial allele frequency of 0.05 and for 16 generations from an initial frequency of 0.20. With an initial frequency of 0.05 there was a serious loss of desired alleles ranging from 0.31 out of 20 for the HS scheme with 24 out of 240 families selected to 9.19 for the S scheme with 6 out of 120 families selected. It was concluded that if as many as 20 cultivars were included in the initial population the selection scheme should be chosen to minimise the loss. With an initial frequency of 0.20 there were no losses with 12 and 24 families selected in the HS and FS schemes respectively, and the highest loss was 2.88 for the S scheme with 6 out of 120 families selected. It was concluded that if as few as five cultivars were included in the initial population a compromise between the initial response to selection and the loss of desired alleles should be sought. Selecting 6, 12 and 24 families for the HS, FS and S schemes respectively, resulted in average relative responses per generation of 2.28, 2.74 and 2.76, respectively for the first five generations, and losses of 0.22, 0.13 and 0.35, respectively after 16 generations. Practical considerations favour the FS scheme over the S scheme.

Key words: Kale breeding – Population improvement – Family selection – Computer simulation – Genetic drift

Introduction

With an outbreeding crop species such as kale (*Brassica oleracea* L.), population improvement by recurrent se-

lection is an important component of any breeding programme. As kale is grown for feeding to cattle and sheep, an important selection criterion is the yield of digestible organic matter in the dry matter (DOMD yield). Single plant selection is impractical because assessment of DOMD yield involves a destructive harvest of the plants. Hence family selection is required so that in addition to a yield trial, selection plots of all the families can be grown to provide plants for seed production.

Plants must pass through a juvenile growth stage and be vernalised before they will flower. Self pollination is largely prevented by a sporophytic incompatibility system, although the breeder can produce selfed seed by bud pollination or special techniques, such as an atmosphere enriched with carbon dioxide (Taylor 1982). The easiest type of family selection scheme for the kale breeder to operate is therefore half-sib family selection on a biennial cycle, in which the half-sib families are produced either by the natural insect pollinators at an isolation site, or by blowflies in special cages or tunnels. Such a scheme has been operated at the Scottish Crop Research Institute (SCRI) since 1971 and has been described by Bradshaw and Mackay (1981). Half-sib matings are also commonly used in forage grass breeding and have been reviewed by Nguyen and Sleper (1983).

The production of full-sib and selfed families of kale involves more work than for half-sib families, but would be worthwhile if the rate of population improvement that could be achieved in practice was faster. As limited resources with a minor crop such as kale usually prevent more than one breeding programme being carried out at a time, decisions on whether or not to change the breeding method must be made on the basis of theoretical models and computer simulation studies. Extensive work has been carried out on recurrent selection methods in maize (*Zea mays* L.) which, unlike kale, is a monoecious, annual, wind-pollinated crop mainly grown for its grain (Hallauer and Miranda 1981). Thus,

the three breeding methods, mass, modified ear-to-row, and S_1 selection, chosen by Choo and Kannenberg (1979a, b) for comparison by computer simulation, as likely from theoretical studies to be the best for maize population improvement, are not the most relevant ones for kale improvement.

This paper therefore reports the results of a computer simulation study of three recurrent selection schemes which are feasible for kale and which involve half-sib, full-sib and selfed families. The effect of type of family, number of families assessed and number of families selected on the loss of desired alleles, as well as on the response to selection, is reported for a range of genetical models.

Materials and methods

Simulation program

The simulation program was written in IMP and run on the Edinburgh Regional Computing Centre's ICL 2988 computer. The methods used were based on those described by Fraser and Burnell (1970). The algorithm of Wichmann and Hill (1982) was used to generate pseudo-random numbers.

Genetical models

In all models there were 20 unlinked loci with two alleles at each. In the additive genetical models the desired homozygote at each locus was assigned a genotypic score of 2, the heterozygote a score of 1 and the undesired homozygote a score of 0. In the dominance models the heterozygote was also assigned a score of 2 (complete dominance).

Four initial populations were considered. Two were in gametic equilibrium with initial average frequencies of the desired allele of 0.05 and 0.20, respectively, although the frequency at any given locus was allowed to deviate by up to one standard deviation from the average. (This constraint prevented extreme frequencies, in particular, zero frequencies.) The other two were disequilibrium models. In one 20 "inbred lines", each with a different locus homozygous for the desired allele and the other 19 loci homozygous for the undesired allele, were random mated for two generations. In the other, 5 "inbred lines", each with a different set of 4 loci homozygous for the desired allele and 16 homozygous for the undesired allele, were also random mated for two generations. Two generations were chosen as the minimum number required to generate genetical variation between families. The random mating was carried out by mating each member of the population to a different member chosen at random. The population size equalled the number of families subsequently produced.

In one set of models the phenotypic score of a family mean equalled the genotypic score whereas in another set it was determined by adding on to the genotypic score an environmental score which was normally distributed with a mean of zero and variance (E) of 2.5 (2.5 is the genetical variance between half-sib families when the frequency of the desired allele is 0.5 at all 20 loci).

Thus, all combinations of additive and dominance models, equilibrium and disequilibrium initial populations, and presence and absence of environmental contributions to the family means made 8 genetical models.

Selection schemes

In all of the schemes, from each plant in the initial population, families of 20 progeny were produced for assessment. The genotypic score of each individual was determined and the family mean calculated. The phenotypic score of the family was then determined. The families were then ranked according to their scores and the best ones chosen. Further progeny were then produced for each of the chosen families in order to form a new population of plants from which to produce the next generation of families to be assessed.

In the half-sib (HS) schemes each plant in the population was mated by 20 plants chosen at random (not necessarily different, but excluding the plant itself) in order to produce 20 half-sib progeny.

In the full-sib (FS) schemes each plant in the population was mated by another plant chosen at random and 20 progeny produced.

In the selfed (S) schemes each plant in the population was selfed to produce 20 progeny in one generation and then in the next generation full-sib families with 20 progeny were produced.

Thus all combinations of type of family, 120 and 240 families assessed and 6, 12 and 24 families selected made 18 selection schemes.

Selection was simulated twice for 20 generations from an initial desired allele frequency of 0.05 and twice for 16 generations from an initial frequency of 0.20, for all combinations of selection schemes and genetical models, making a total of 576 runs. (After these numbers of generations alleles had usually either been lost or had reached a relatively high frequency. There were some cases of just one locus with an allele frequency below the starting frequency, but loss of these in further generations would not alter the conclusions.)

Analyses

The number of desired alleles lost, out of 20, by generation 20 from a starting frequency of 0.05, and by generation 16 from a starting frequency of 0.20, was counted for each simulation.

The initial relative responses to selection per generation were assessed as the difference, divided by 5, between the population mean after five generations and the initial population mean.

The numbers of alleles lost and the responses to selection were analysed by analyses of variance with both the selection schemes and the genetical models as fixed effects.

Results

The analyses of variance are shown in Table 1. The differences between genetical models, between selection schemes, and their interaction were all statistically significant when tested against the pooled mean square for duplicate runs, for all four sets of data. As examination of the tables of means showed that the interactions did not involve major changes in ranking, the means for selection schemes and genetical models given in Tables 2–5 are adequate for discussing the main conclusions.

The analyses of variance were taken further by partitioning the 17 degrees of freedom for selection schemes into the three factors shown in Tables 2a to 5a and the 7 degrees of freedom for genetical models into the

Table 1. Analyses of variance

Source of variation	df	Mean squares			
		Loss 0.05	Loss 0.20	Response 0.05	Response 0.20
Genetical models	7	80.476***	10.134***	14.793***	6.117***
Selection schemes	17	111.440***	10.256***	3.482***	5.134***
Models × schemes	119	2.628***	1.140***	0.082***	0.096***
Duplicate runs	144	1.819	0.531	0.050	0.049

*** $P < 0.001$

Table 2. Number of alleles out of 20 lost during 20 generations of selection from an initial frequency of 0.05

a Selection schemes

	6	6	12	12	24	24	
Families selected	6	6	12	12	24	24	
Families assessed	120	240	120	240	120	240	
HS	4.69	4.00	2.06	1.44	1.50	0.31	2.33
FS	5.81	5.81	3.88	2.81	2.38	0.81	3.58
S	9.19	8.94	7.00	6.25	3.63	2.75	6.29
	6.56	6.25	4.31	3.50	2.50	1.29	4.07

Average SE of items in centre of a. 0.337

b Genetical models

	Additive		Dominance		
	Equilibrium	Disequilibrium	Equilibrium	Disequilibrium	
E=0	3.56	3.94	2.06	3.31	3.22
E=2.5	5.28	6.94	3.11	4.36	4.92
	4.42	5.44	2.58	3.83	4.07

Average SE of items in centre of b. 0.225

Table 3. Number of alleles out of 20 lost during 16 generations of selection from an initial frequency of 0.20

a Selection schemes

	6	6	12	12	24	24	
Families selected	6	6	12	12	24	24	
Families assessed	120	240	120	240	120	240	
HS	0.38	0.06	0.00	0.00	0.00	0.00	0.07
FS	0.81	0.69	0.06	0.19	0.00	0.00	0.29
S	2.88	2.13	1.06	0.88	0.44	0.25	1.27
	1.35	0.96	0.38	0.35	0.15	0.08	0.55

Average SE of items in centre of a. 0.182

b Genetical models

	Additive		Dominance		
	Equilibrium	Disequilibrium	Equilibrium	Disequilibrium	
E=0	0.39	0.58	0.03	0.25	0.31
E=2.5	1.08	1.58	0.14	0.31	0.78
	0.74	1.08	0.08	0.28	0.55

Average SE of items in centre of b. 0.121

Table 4. Response to selection per generation from an initial allele frequency of 0.05: average of first 5 generations

a Selection schemes							
Families selected	5%	2½%	10%	5%	20%	10%	
Families selected	6	6	12	12	24	24	
Families assessed	120	240	120	240	120	240	
HS	1.55	1.83	1.34	1.61	1.04	1.38	1.46
FS	2.21	2.53	1.84	2.41	1.51	2.05	2.09
S	2.28	2.65	2.02	2.52	1.83	2.28	2.26
	2.01	2.34	1.73	2.18	1.46	1.90	1.94
Average SE of items in centre of a. 0.0558							
b Genetical models							
	Additive			Dominance			
	Equilibrium	Disequilibrium		Equilibrium	Disequilibrium		
E=0	2.01	1.67		2.88	2.47		2.26
E=2.5	1.33	0.90		2.37	1.88		1.62
	1.67	1.29		2.62	2.17		1.94
Average SE of items in centre of b. 0.0372							

Table 5. Response to selection per generation from an initial frequency of 0.20: average of first 5 generations

a Selection schemes							
Families selected	5%	2½%	10%	5%	20%	10%	
Families selected	6	6	12	12	24	24	
Families assessed	120	240	120	240	120	240	
HS	2.11	2.45	1.94	2.32	1.51	1.96	2.05
FS	2.81	3.15	2.54	2.93	2.06	2.61	2.68
S	3.15	3.60	2.83	3.42	2.41	3.11	3.09
	2.69	3.07	2.44	2.89	1.99	2.56	2.61
Average SE of items in centre of a. 0.0554							
b Genetical models							
	Additive			Dominance			
	Equilibrium	Disequilibrium		Equilibrium	Disequilibrium		
E=0	2.83	2.52		3.16	2.89		2.85
E=2.5	2.27	1.86		2.86	2.47		2.37
	2.55	2.19		3.01	2.68		2.61
Average SE of items in centre of b. 0.0369							

three factors shown in Tables 2 b–5 b. However, the results are not presented because of their complexity. Not only were the six factors statistically significant (with the exception of the number of families assessed for the number of alleles lost from a starting frequency of 0.20), but so were a number of first order and higher order interactions.

Discussion

Selection schemes

The selection schemes were chosen as feasible with kale. If the plants in the selection plots are overwintered and

transplanted for seed production in the spring, all three schemes involve a biennial cycle. The HS scheme involves the least work as natural insect pollination is involved, and it is relatively easy to produce 240 families with enough seed (> 20 g) for assessment in four relatively large plots (> 15 m²). The FS scheme involves more work as plants would need to be planted in pairs, covered with insect proof bags, and blowflies introduced. Again plenty of seed would be produced. The S scheme requires even more work as it involves overcoming the incompatibility system, and seed could be in short supply, necessitating smaller and fewer plots, and hence a larger environmental component to the variance of family means.

If a greenhouse with heating and lighting were available for "out of season" pollinating in winter, all three schemes could be carried out on an annual cycle. However, this would involve hand pollinations which would limit both the number of families and quantity of seed produced. Under these circumstances 120 families assessed might be more realistic than 240, with as few as 20 progeny per family.

As plant breeders today are under pressure to produce improved cultivars as quickly as possible, although factors affecting the rate of response to selection were varied and the responses to selection over the first five generations analysed, all of the schemes incorporated a number of features aimed at rapid improvement. Only schemes in which 20% or less of the families assessed were selected were considered. Plants from the selected families were recombined to produce new families for assessment without additional cycles of recombination. In the disequilibrium populations selection was started as soon as genetical variation existed between families.

For ease of simulation, the actual simulations departed from practice in a number of respects. Only 20 progeny of each family were assessed whereas the relatively large plots envisaged in the biennial cycles would contain well over 100 plants. In the half-sib scheme independent samples of pollen parents were taken for assessment and for producing the next generation whereas they would be the same in practice. In the full-sib scheme the second parents were chosen at random from the population used as first parents, whereas in the biennial cycles twice as many plants as families required would be selected and paired. However, it is not thought that these differences alter the main conclusions to be drawn from the results.

Genetical models

Some discussion of the genetical models chosen is required as they inevitably involved oversimplifications. In kale breeding it is likely that the initial population would be constructed by intercrossing cultivars thought to complement one another for desired alleles, and hence for simplicity simulated by intercrossing "inbred lines". The number of such cultivars would probably be in the range of five to twenty. Hence the disequilibrium models are more realistic than the simpler equilibrium models, which were also included as they are often used in theoretical models. The environmental variance of 2.5 was chosen on the basis of the results of the HS scheme reported by Bradshaw and Mackay (1981). As a contrast, and an indication of what would happen if the heritability could be increased by reducing the environmental component of the variance of family means, the case of no environmental variance was simulated. For yield in kale the directional dominance model is more realistic than the simpler additive model as inbreeding depression is known to occur. Again though, the additive model was included as a contrast and may be relevant to other characteristics of importance in kale breeding.

The choice of 20 unlinked loci with genes of equal effect was both convenient and arbitrary as the actual number of loci segregating, and the size of the gene effects, in actual kale

breeding populations is unknown. Although this choice was almost certainly an oversimplification, it was thought adequate for the purpose of the investigation.

Loss of alleles

Before discussing their implications, it is worth considering the results themselves. Initial allele frequency had a large effect on the loss of desired alleles. The average loss fell from 4.07 to 0.55 as the initial frequency increased from 0.05 to 0.20. Choo and Kannenberg (1979 b) found losses of desirable alleles under all their selection methods when the initial frequency was 0.10 but no losses when it was 0.50, and Wright (1980) did not obtain any losses from an initial frequency of 0.30 in his simulation of S_1 family selection. The number of families selected also had a large effect with the loss rising from 1.90 to 6.41 as the number selected was reduced from 24 to 6 with an initial frequency of 0.05, and from 0.12 to 1.16 when it was 0.20. The number of families assessed had less effect although losses were fewer for 240 families assessed than for 120: i.e. doubling the family size helped to reduce losses. Choo and Kannenberg (1979 b) also found losses were greater under higher selection intensity (5%) which meant that in their S_1 family selection scheme 20 out of 400 families were selected. The loss of alleles was greatest with the S scheme and least with the HS scheme, as would be expected from considering effective population sizes. Although not directly comparable with the S scheme, Choo and Kannenberg (1979 b) found that more alleles were lost with S_1 selection than with modified ear-to-row or mass selection.

The presence of environmental variation resulted in a greater loss of desired alleles, presumably because the alleles remained at a relatively low frequency for longer. The loss was also greater under the additive genetical model than under the directional dominance model, presumably because under an additive model two heterozygous loci receive the same genotypic score as two unlike homozygous loci. Choo and Kannenberg (1979 b) also found that more alleles were lost under an additive model but found that two simulation runs were too few to determine the effect of their heritabilities on genetic drift. The disequilibrium models also resulted in greater loss than the equilibrium models. However, it must be remembered that for the equilibrium models allele frequency was constrained to within one standard deviation of the mean in the initial population, and for the disequilibrium models the initial population resulted from two generations of random mating in order to generate genetical variation between families. This led to more variation in the desired allele frequency between loci in the disequilibrium populations, although there were no losses before selection commenced. In some preliminary simulation runs it was demonstrated

that further generations of random mating did result in loss of desired alleles so that it was important to start selection as soon as possible in the disequilibrium populations.

Response to selection

The ranking of the responses for both initial allele frequencies were as expected from simple theoretical considerations. The S scheme was best and the HS scheme worst. Six out of 240 families selected was best and 24 out of 120 was worst. The presence of environmental variation reduced the response. The additive and dominance models are not directly comparable as both the additive genetic variance and the total genetic variance were larger in the dominance models, as a result of the assigned genotypic scores, and hence the responses were greater. The response was larger in the equilibrium populations than in the disequilibrium populations. Examination of the summary statistics for each simulation revealed that this was because there was initially less genetical variation between families in the disequilibrium populations. It is also worth pointing out that it was impossible for all of the desired alleles to exist in a single individual in the initial disequilibrium populations.

The initial responses to selection were less from an initial allele frequency of 0.05 than from a frequency of 0.20 because there was less genetical variation at the low frequency. These differences also emphasise the fact that the responses were not linear over the complete range of generations.

Implications for breeding programmes

Finally it is necessary to consider the practical implications of the results. When only the initial response (first five generations) to selection is considered the S scheme was superior to the FS scheme which in turn was superior to the HS scheme, and selecting 6 out of 240 families, the highest selection intensity considered, was best. However, as indicated earlier in the discussion, when operating on a biennial cycle the selfed families could have a larger environmental component to the variance of their family means as a result of smaller and fewer plots, and this would reduce their relative response to selection.

For selection programmes continuing for rather more than five generations the loss of desired alleles is also an important consideration. If the breeder wishes to include as many as 20 cultivars in his initial population, then some loss of desired alleles is likely to occur, but can be kept to a minimum by careful choice of selection scheme. Thus the HS scheme was superior to both the FS and S schemes and selecting fewer than 24 families, out of 240 assessed, would seem undesirable.

As a result of their simulations of S_1 family selection Choo and Kannenberg (1979b) advocated that at least 20 S_1 families should be recombined each cycle. The empirical advice given to maize breeders by Hallauer and Miranda (1981) in their breeding plans for a number of family selection schemes was to recombine 20 to 30 families. Whilst the results presented here for kale are consistent with this advice, it can be concluded that the numbers should be varied according to the type of family.

When as few as five cultivars were included in the initial population there were only slight losses when 6, 12 and 24 families were selected in the HS, FS and S schemes, respectively, and no losses with 12 and 24 families in the HS and FS schemes, respectively. Thus a compromise between the initial response to selection and the loss of desired alleles may be sought. For example, selecting 6, 12 and 24 families for the HS, FS and S schemes, respectively might be acceptable. For 120 families assessed the respective average responses were 2.11, 2.54 and 2.41 and for 240 families assessed they were 2.45, 2.93 and 3.11. Thus there was little to choose between the FS and S scheme, although practical considerations would favour the FS scheme over the S scheme as bud pollinations would be avoided. Although both schemes were superior to the HS scheme, when working on a biennial cycle the HS scheme of Bradshaw and Mackay (1981), involving natural insect pollination, could remain both attractive and competitive if its ease of operation resulted in more families being assessed.

References

- Bradshaw JE, Mackay GR (1981) Kale population improvement and cultivar production: In: Gallais A (ed) Quantitative genetics and breeding methods. Eucarpia Proc 4th Meeting Sec Biomet Plant Breed, pp 71–76
- Choo TM, Kannenberg LW (1979a) Relative efficiencies of population improvement methods in corn: a simulation study. *Crop Sci* 19:179–185
- Choo TM, Kannenberg LW (1979b) Changes in gene frequency during mass, modified ear-to-row, and S_1 selection: a simulation study. *Crop Sci* 19:503–509
- Fraser A, Burnell D (1970) Computer models in genetics. McGraw-Hill, New York
- Hallauer AR, Miranda JB (1981) Quantitative genetics in maize breeding. Iowa State University Press, Ames Iowa
- Nguyen HT, Slepner DA (1983) Theory and application of half-sib matings in forage grass breeding. *Theor Appl Genet* 64:187–196
- Taylor JP (1982) Carbon dioxide treatment as an effective aid to the production of selfed seed in kale and Brussels sprouts. *Euphytica* 31:957–964
- Wichmann BA, Hill ID (1982) An efficient and portable pseudo-random number generator. *Appl Stat* 31:188–190
- Wright AJ (1980) The expected efficiencies of half-sib, testcross and S_1 progeny testing methods in single population improvement. *Heredity* 45:361–376