Theor Appl Genet (1991) 82:363–367

THEORETICAL
AND
APPLIED GENETICS

Selection for local adaptation in a plant breeding programme

N. W. Simmonds
Edinburgh School of Agriculture, West Mains Road, Edinburgh EH9 3JG, Scotland

Received December 4, 1990; Accepted February 20, 1991
Communicated by J. W. Snape

Summary. Regressions of yields of cultivars upon means of sets of cultivars over diverse environments are often used as measures of stability/adaptability. Prolonged selection for performance in environments of high yield potential has generally led to unconscious selection for high regressions. If adaptation to poor environments is required (as it often is in Third World agriculture), common sense suggests that low regressions could be exploited for the purpose. Simulations show that systematic selection in the poor environment is required, not merely trials of potential cultivars after selection in a good environment. In effect, systematic exploitation of a GE interactions effect is proposed. The effects are large enough to reduce correlated responses in different environments to zero. Orderly experimental studies are needed but not available. What information there is does not disagree with the theory developed here.

Key words: Adaptation – Varietal stability – GE interactions – Simulation

Interaction

All plant breeding programmes are beset by the problems of genotype × environment interaction (GE) effects. Although there is a general awareness of the problems, little is done in practice beyond testing potential new cultivars in diverse environments in the hope of revealing wide adaptation (i.e., stability) of performance as well as good mean performance. Given that many programes are locally directed towards adaptation to agro-ecological environments similar to that of the breeding station, this may often be a reasonable procedure.

As an aid to recognizing range of adaptation, regressions of variety performance on the means of other varieties in the same trials have been widely used. Their use goes back to Yates and Cochran (1938), followed by Finlay and Wilkinson (1963), later widely popularized by Eberhart and Russell (1966). In all, there must have been hundreds of publications on the subject. Generally, it is argued that varieties with low regressions are stable, unresponsive and vary little, while those with high slopes are responsive, more variable and especially adapted to high-yielding environments. Despite the huge volume of publications on the subject, it is not clear that actual decision making has been appreciably influenced. Rather, selection for high performance in ‘high’ environments has tended to generate more or less unconscious selection of new varieties with high regressions relative to other candidates (Simmonds 1979, 1981). Hence, there is a question that has rarely been asked, let alone experimentally tested, namely: is there potential merit in deliberately selecting for good performance in low environments, in the expectation of correlated responses towards low regression coefficients? In short, can the regressions be exploited rather than merely observed? This paper gives a simple, perhaps even overly simple, exploration of that question by means of numerical simulation. Conclusions are especially relevant to plant breeding programmes in Third World countries.

Model

The main features are shown in Fig. 1. I assumed a starting population of 5,000 genetic entities (lines at or near fixation, clones) with a mean yield $Y_M = 10$ in a middling environment (E_M) and means in contrasted low and high environments (E_L, E_H) of $Y_L = 5$, $Y_H = 15$. I assumed a constant genetic standard deviation of yield ($\sigma_{GY} = 0.5$) in all environments and a variable regression with mean $b = 1.00$, normally distributed

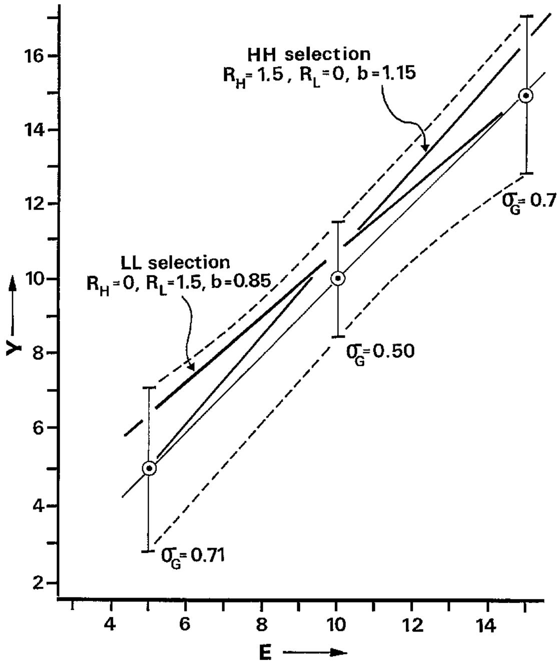

Fig. 1. Regression scheme adopted (see text). Y, yield; E, environment; HH LL, high and low selection patterns; R, genetic advances; *b*, regression coefficients; G, genetic standard deviations. Responses are rounded versions of data in Table 1. Correlated responses in the contrasted environment are zero

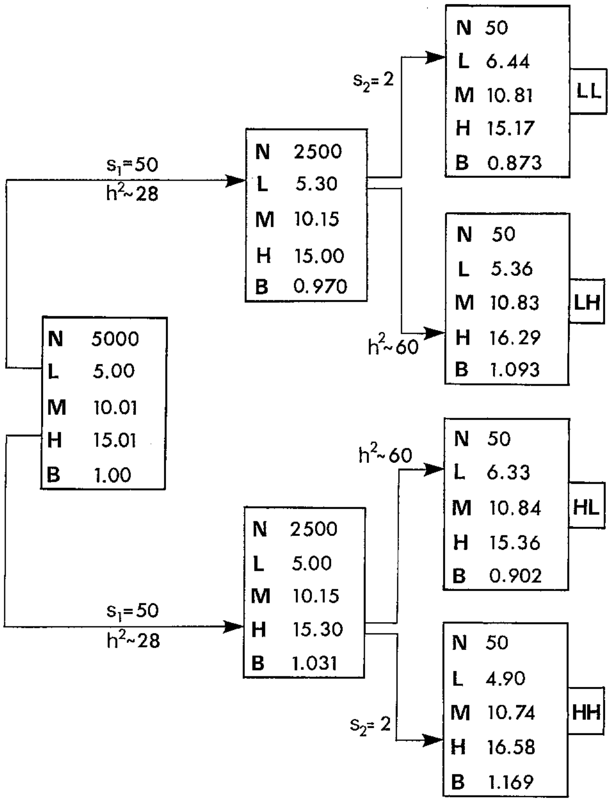

Fig. 2. Example of results of selection patterns examined. *N*: numbers of entries; L, M, H: low, medium, high; *B*: regression coefficient; S_1 and S_2: first and second cycle rates of selection; h^2: heritability. Four selection patterns are displayed (LL, LH, HL, HH) and results in the *top* and *bottom* right-hand boxes (LL and HH) correspond with the means given in Table 1

around a central mean at $Y_M=10$ with genetic standard deviation $\sigma_{GB}=0.10$. Thus, in E_M, the regression generated no deviation, but in E_L and E_H it produced deviations in Y of $\pm 5b$. Since $\sigma_{GB}=0.10$, a 'reasonable' genetic range for *b* was 0.80 to 1.20, in accord with experience. The general result, which is intuitively obvious, is apparent in Fig. 1. Selection for Y in E_M had little scope for success because it could not exploit the regression; in E_H and E_L, the regressions effectively enhanced genetic variance, permitting stronger responses in Y, with correlated responses in *b*, upwards to >1.0 in E_H, downwards to <1.0 in E_L. Expected genetic variances in E_L and E_H are the sums of two components, one (σ^2_{GY}) free of regression, the other dependent upon it; thus: $(0.5^2+5^2\cdot\sigma^2_{GB})=(0.25+0.25)=0.5$. Hence (see Fig. 1), σ_{GT} (total genetic σ) rises from 0.5 at E_M to $\sqrt{0.5}=0.71$ at the extremes.

For the purpose of this model, I assumed a two-stage selection process with errors of yield and regression coefficient as follows: $\sigma_{EY}=1.0$ and $\sigma_{EB}=0.10$ in the first stage, and $\sigma_{EY}=0.50$, $\sigma_{EB}=0.05$ in the second stage. Thus, heritability was enhanced in stage 2 by reason of reduced errors. Generally, h^2 was about 28% in stage 1 and about 60% in stage 2 (Fig. 2), independent of the precise selection pattern chosen.

I assumed six combinations of selection rates, S_1 and S_2, all giving an outcome of $S_1S_2=1\%$. The final product was therefore a constant 50 selections. Selection rates (percent) examined were: 80/1.25, 50/2, 30/3.33, 20/5, 10/10, 5/20.

To describe outcomes, I used the following notation. *R* is genetic advance in *Y*; for example, in Fig. 1, $R_L=1.5$ denotes an advance of 1.5 units over the starting mean, from 5.0 to $Y_L=6.5$. Selection regimes (Fig. 2) are denoted by two letters indicating the environments of the two cycles in sequence, thus: HH, HL, LH, LL. *M* is introduced where necessary. Nearly all interest centres on HH and LL. Where needed, a test environment is indicated after an oblique stroke; thus HH/*L* denotes two cycles of selection in E_H followed by evaluation in E_L.

Results

The six different patterns of selection detailed above (from $S_1=80$ to $S_1=5$) gave nearly identical results in respect of LL and HH selection. There were some differences for LH and HL, but these are of no great interest and need not be further considered. LH and HL results are illustrated in Fig. 2, but attention is otherwise concentrated hereafter on LL and HH. The main result, with slightly rounded and simplified figures, is given in Fig. 1 and data are summarised in Table 1. LL selection made good progress in E_L but was accompanied by a strong correlated response for low *b*, and hence selections made near-zero progress in E_H. The complementary selection, in E_H, behaved symmetrically, with a correlated increase in *b*. Neither selection made substantial progress in E_M. (Several simulations of selection in E_M made rather better progress as would, of course, be predicted but this is not an interesting result.)

A typical individual run is shown in Fig. 2. Results for LH and HL are, as expected, intermediate in all parameters. Neither made good average progress, although se-

Table 1. Simulations of responses to selection

		LL selection			HH selection		
		LL/L	LL/H	b	HH/H	HH/L	b
A. Whole populations	(1)	6.48 ± 0.013	15.02 ± 0.028	0.854 ± 0.003	16.51 ± 0.018	4.93 ± 0.015	1.158 ± 0.002
($n = 50$)	(2)	7.30	16.22	0.707	17.26	6.22	1.317
B. Split populations	(1)	6.50 ± 0.052	14.92 ± 0.062	0.842 ± 0.010	16.49 ± 0.038	4.92 ± 0.056	1.157 ± 0.006
($n = 25$)	(2)	7.11	15.77	0.742	17.08	5.99	1.284

Note: A refers to whole populations, starting at $n = 5{,}000$, pooled over six patterns of selection (see text). B refers to populations of 5,000 split into two contrasted halves, one HH and other LL, at $S_1 = 50$, $S_2 = 2\%$, so that surviving samples were $n = 25$. In each part, (1) gives means for the whole populations (with $n = 50$, 25 respectively) while (2) gives estimates of the best five selections based on a table of normal ordinal scores

lected samples (all $n = 50$) usually did contain a few entries more or less excellent in E_L and E_H. The second cycle of selection (L in HL, H in LH) was the most potent, which is not surprising because heritability was constructed to be higher. In general, LH and HL selection had little to offer for good adaptation to any environment. If practised, it would be best to aim at intense selection in the second cycle in the environment for which adaptation is sought (Fig. 2).

So far the result is clear: select in the environment to which adaptation is sought. Correlated responses in respect of performance in contrasted environments support this conclusion. Correlations between Y_L and Y_H after LL and HH selection in populations of $n = 50$ were in the range $r = -0.27$ to $+0.21$, with mean $r = -0.07$. In effect, therefore, correlations were zero and performance in neither environment predicted performance in the other. Correlations of Y with b are also of some interest; thus:

Selection	Test	
	L	H
LL	-0.56	$+0.83$
HH	-0.88	$+0.59$

Low regressions favour adaptation to E_L, high to E_H, and vice versa. Thus, there is no single relationship between yield and regression; any relation will depend upon the test environment and history of selection.

Since it was clear that alternating selection (LH, HL) offered little attraction for joint adaptation to both environments, it was obviously worth considering the strategy of splitting the population into two halves (outcomes $n = 25$ HH + 25 LL). Results are shown in Table 1. Mean responses are virtually unchanged, although the smaller sets of selections would leave the breeder less room for manoeuvre. Furthermore [lines (2) in Table 1], extreme selections would probably not be so good. However, it is doubtful wheather it is reasonable to assign magnitudes to quantities in the tails of distributions, whether normal or, more probably, nonnormal. These lines (2) merely show that sample sets of 25–50 selections should still leave room for some further advance and serve to reemphasise the conclusion stated above: select in LL and/or HH but expect poor mean results from LH and HL.

The two contrasted selection schedules, HH and LL, generated samples with very different mean b (Table 1). Means and mean standard deviations within samples were:

$$\text{LL} \quad 0.854 \pm 0.087\,, \qquad \text{HH} \quad 1.158 \pm 0.094\,.$$

Thus, confidence limits (at $\pm 2\sigma$) for the two would be about 0.68–1.03, 0.97–1.35. At least a little overlap between every pair of samples would be expected and is found. These regressions all refer to the original population as base defining $b = 1$. After selection, sets of LL- and HH-adapted entities tested in isolation would set new standards for the base populations defining $b = 1$. Thus, after rescaling, the mean b of each group becomes unity, with standard deviations and confidence ranges: LL 0.104, 0.79–1.21 and HH 0.162, 0.84–1.16. As was mentioned above and will be reiterated later, these figures lie well within the range of experience.

Discussion

There are eight points to make. The first is that the main result is, in a sense, obvious. If the correlation between performances in E_L and E_H is less than perfect, then direct selection should always be best. I am obliged to referees for calling attention to this, but observe that what is obvious to biometrical geneticists may not be so obvious to plant breeders. In practice, plant breeders regularly select in E_H and hope for performance in E_L, even when they know that the procedure is suboptimal. I am obliged, also, to referees for pointing out that my models are overly simple; so they probably are, but one has to start somewhere and the best possible outcome would be, not more or more refined models, but good experiments.

Second, I assumed that the regression attaching to a genotype was moderately heritable/repeatable. This assumption is, of course, inherent in all uses of the regressions and much general experience favours the idea. Explicit evidence is not abundant, but some does exist: Eberhart and Russell (1966) showed that there was a large GCA component for b among maize inbreds; Patterson's (1980) wheat data showed that the b were moderately repeatable over years; Walker's and Simmonds's (1981) sugarcane data gave a (weak) correlation between the b estimated in trials and those on-farm. It would be good to have more extensive studies but the fact of fair repeatability seems certain. The b, as far is known, are all, for practical purposes, linear; I have seen no suggestion of curved relationships and, given the usual presence of 'noise', it would be hard to establish any consistent nonlinearity.

Third, the size of the b assumed for this study was in an approximate genetic range of 0.8 to 1.2. This accords well with experience (recalling that observations will tend to be stretched by errors). Thus, consider the following examples, only a few from what could be a very long list: 0.8 to 1.2 (barley, Ceccarelli 1989); 0.7 to 1.3 (maize, Eberhart and Russell 1966); 0.7 to 1.3 (oats, Langer et al. 1979); 0.8 to 1.3 (wheat, Patterson 1980); 0.5 to 1.4 (wheat, Brennan and Byth 1979); 0.7 to 1.2 (wheat, Laing and Fischer 1977; see also Simmonds 1979); 0.8 to 1.7 (sugarcane, Simmonds 1979); 0.8 to 1.2 and 0.7 to 1.4 (potatoes, Simmonds 1980); 0.6 to 1.7 and 0.4 to 1.3 (sugarcane, Walker and Simmonds 1981). In short, a substantial range of b is normal and the values here assumed for modelling are, if anything, conservative.

A relevant, though hardly critical, point about the regressions concerns their relation to yield. That there must tend to be some relation is obvious from Fig. 1. Many authors (e.g. Eberhart and Russell 1966; Patterson 1980; Langer et al. 1979; Laing and Fischer 1977; Brennan and Byth 1979) have detected positive correlations between Y and b, while others have found none. As the simulations show, however, there is no general expectation; negative correlations are likely in trials done in E_L, but the same genotypes could yield a positive correlation if tested in E_H. Since trials are usually done on experiment stations, a bias towards positive correlations is indeed expected. Negative ones are to be expected only when someone does the right experiments.

Fourth, selection ab initio in strongly contrasted sites (E_L, E_H) is potent in generating local adaptation, to which variation in terms of regression (on the model chosen) makes a major contribution. Obviously, choice of other parameters could either strengthen or weaken this conclusion. Thus, choice of a more 'noisy', less repeatable b and higher genetic variance independent of b would weaken it. Very diverse choices are possible but, for want of well-estimated parameters, more simulation would not be profitable. Response is plainly maximised by consistent selection (i.e. LL and HH). The opposite, inconsistent selection (LH, HL), gives weak responses at the limits and tends not to exploit the regressions effectively (Fig. 2). If adaptation to two contrasted environments is being considered, it would be better to divide the programme into two parts ab initio (LL and HH), rather than the hope that mixed-site compromises would be effective.

Fifth, there are a good many reports of trials carried out in contrasting environments, usually with a view to choosing trial sites. However, only rarely, it seems, has the idea of systematic *selection* in contrasted environments even been contemplated. The works of Frey (1964), Srivastava et al. (1983), Sage et al. (1984), Ceccarelli (1987), Ceccarelli and Grando (1989, 1990) and Atlin and Frey (1989) are relevant. They are all concerned with small-grain cereals (barley, wheat, oats). The general trend of conclusion is that adaptation to an environment (whether E_L or E_H) is best achieved by selecting in that environment, as argued here, but this stands in contrast to much plant breeding practice, which selects on the home station and hopes for adaptation elsewhere. The sites used by Sage et al. (1984) were not strongly contrasted and only weak hints of local adaptive response emerged, but the other cases were all rather clear. By far the clearest example of selection in contrasted environments comes from South African sugarcane (Simmonds 1984). The experiment was based upon a population of seedlings sampled and then selected in two highly contrasted environments, the clonal products being finally tested in those same environments. A strong GE component of response for sugar yield emerged, with regressions roughly estimated as 0.7 for selections from E_L and 1.3 for selections from E_H. It would be beneficial to have many more orderly, orthogonal experiments of this kind for diverse crops. Some, indeed, are in progress (S. Ceccarelli, personal communication).

Sixth, traditionally, most plant breeding takes place in E_H on the experiment station and, as I remarked above, this may be quite reasonable in countries that practise high-input agriculture in which farms and research stations employ very similar husbandry. (Even so, experiment station yields are often substantially higher than farm yields – examples in Simmonds 1980.) In Third World countries, matters are often different, with experiment stations getting good yields at high inputs, but with breeding programmes aimed at the needs of very diverse farmers. This is broadly true of the Consultative Group (CG) system of International Agricultural Research Centers (Anderson et al. 1988). Some of them (e.g. CIMMYT on wheats, IRRI on rices) have had excellent successes with new varieties aimed at high-input farmers, precisely on the rich, temperate country model; indeed, this exploitation of a positive GE-interaction-based package was the core of the success of the Green Revolution (Simmonds 1979, 1981; Anderson et al. 1988; Lipton and Longhurst 1989). In this sort of context, selection has inevitably, but unconsciously, been for the high yielding, high-b variety, even though (following Eberhart and Russell 1966) $b=1$ is theoretically preferred as showing 'stable', 'non-sensitive' performance. However, the great majority of tropical small farmers gets low yields at low levels of inputs in environments that are inherently hard, inhospitable. Whatever successes may fairly be attributed to the Green Revolution, the CG system can claim only local/minor successes for the multitude of low-input food crop enterprises that underlie much tropical agriculture. The appropriate response to this situation constitutes the next point.

Seventh, in the light of the literature cited above and simulations reported here, the sensible response by plant breeders seeking to breed for E_L would be to select in E_L; to *select*, be it noted, not merely do trials after selecting in E_H. There would be difficulties, to be sure. Reliable/repeatable low environments might not be easily found or managed (but good environments are not always all that repeatable either). It is sometimes said (e.g. Frey 1964; Rosielle and Hamblin 1981; Roy and Murty 1970) that yield heritability would be lower in E_L, but the data offered are hardly convincing. Cecarrelli (1987, 1989), working with barley in Syria, found no difference and I have seen no convincing a priori arguments. Indeed (as in my simulations), variable regressions might actually enhance heritabilities towards the extremes. In any event, it is certain that the rather simple idea of explicitly seeking to build adaptation to E_L and therefore to exploit the regressions is not often even recognised, let alone practised.

Eighth, animal breeders have long regarded adaptation to contrasted environments as reflecting correlated responses, but usually with a view towards maximal mean performance (Falconer 1990). Correlated responses do appear, albeit rarely, in plant breeding literature (e.g. in Rosielle and Hamblin 1981; Atlin and Frey 1989). In general, plant and animal breeders share a more or less common view of the regressions as reflecting sensitivity/stability/adaptability/reliability of performance, rather than as a feature to be practically exploited. However, Falconer's (1990) Fig. 2 on mice selected for growth at different planes of nutrition shows a pattern remarkably like that of Fig. 1 herein and low (but not zero) correlated responses; the mouse breeder would best develop E_L adaptation by selecting in E_L.

References

Anderson JR, Herdt RW, Scobie GM (1988) Science and food. The CGIAR and its partners. World Bank, Washington, DC

Atlin GN, Frey KJ (1989) Predicting the relative effectiveness of direct versus indirect selection for oat yields in three types of stress environments. Euphytica 44:137–142

Brennan PS, Byth DE (1979) Genotype × environmental interactions for wheat yields and selection for widely adapted wheat genotypes. Aust J Agric Res 30:221–232

Ceccarelli S (1987) Yield potential and drought tolerance of segregating barley populations in contrasting environments. Euphytica 36:265–274

Ceccarelli S (1989) Wide adaptation: how wide? Euphytica 40:197–205

Ceccarelli S, Grando S (1989) Efficiency of empirical selection under stress conditions in barley. J Genet Breed 43:25–31

Ceccarelli S, Grando S (1990) Environment of selection and type of germplasm in barley breeding for stress conditions J Genet Breed (in press)

Eberhart SA, Russell WA (1966) Stability parameters for comparing yields. Crop Sci 6:36–40

Falconer DS (1990) Selection in different environments: effect on environmental sensitivity and on mean performance. Genet Res Camb 56:57–70

Finlay KW, Wilkinson GN (1963) The analysis of adaptation in a plant breeding programme. Aust J Agric Res 14:742–754

Frey KJ (1964) Adaptation of oat strains selected under stress and non-stress environmental conditions. Crop Sci 4:55–58

Laing DR, Fischer RA (1977) Adaptation of semi-dwarf wheat cultivars to rainfed conditions. Euphytica 26:129–139

Langer I, Frey KJ, Bailey T (1979) Associations among productivity, production response and stability indexes in oat varieties. Euphytica 28:17–24

Lipton M, Longhurst R (1989) New seeds and poor people. Unwin Hyman, London

Patterson HD (1980) Yield sensitivity and straw shortness in varieties of winter wheat. J Nat Inst Agric Bot 15:198–204

Rosielle AA, Hamblin J (1981) Theoretical aspects of selection for yield in stress and non-stress environments. Crop Sci 21:943–946

Roy NN, Murty OR (1970) A selection procedure in wheat for stress environments. Euphytica 19:509–521

Sage GCM, Roffey AP, Stanca AM (1984) Simultaneous selection of segregating two-row winter barley material in England and Italy. Euphytica 33:187–198

Simmonds NW (1979) Principles of crop improvement. Longman, London

Simmonds NW (1980) Comparisons of the yields of four potato varieties in trials and in agriculture. Expl Agric 16:393–398

Simmonds NW (1981) Genotype (G) environment (E) and GE components of crop yields. Exp Agric 17:355–362

Simmonds NW (1984) Decentralized selection. Sugar Cane 6:8–10

Srivastava JP, Jana S, Gautam PL, Niks RE (1983) Parallel selection: an approach to increase yield and stability. Proc 6th Int Wheat Genet Symp, Kyoto, pp 725–733

Walker DIT, Simmonds NW (1981) Comparisons of the performance of sugar-cane varieties in trials and agriculture. Exp Agric 17:137–144

Yates F, Cochran WG (1938) The analysis of groups of experiments. J Agric Sci Camb 28:556–580