DOES SUBDIVISION OF THE SUNFLOWER GROWING REGION OF ARGENTINA INCREASE THE RESPONSE TO SELECTION?

Abelardo J. de la Vega, Advanta Semillas S.A.I.C. Ruta Nac. 33 Km 636, CC 559,
2600 Venado Tuerto, Argentina
E-mail: avega@waycom.com.ar

Abstract

The central and northern regions of Argentina are different mega-environments for sunflower, and breeding programs are able to exploit specific adaptation to each region to increase gains from selection. Division of a crop region could result in an increase in the genetic variance ($\sigma^2 g$) and the $\sigma^2 g$ to genotype × environment interaction variance ($\sigma^2 ge$) ratio within-subregion, but implies the division of testing resources, which could result in loss of precision in the estimation of genotype means. A correlated response to the selection model was applied to a trial data set consisting of 10 sunflower hybrids grown over 46 environments in Argentina. The predicted ratio of correlated response (*CR*) in a subregion to indirect selection in the target region relative to direct response (*DR*) in the subregion was used to determine if division of the sunflower region of Argentina is likely to increase response to selection. *CR* and *DR* depend on the genetic correlation between the target region and its subregions (r_g) and the repeatability in each environment. Subdivision of the Argentine sunflower region in northern and central subregions strongly increased the $\sigma^2 g$ to total $\sigma^2 ge$ ratios within-subregion. Division of the testing resources did not result in lower repeatabilities within-subregion. This, together with a low *rg* observed (0.36), resulted in *CR*/*DR* ratios of 0.31 and 0.35 for the central and northern subregions, respectively. These estimates indicate that dividing the sunflower region of Argentina into central and northern subregions and exploiting specific adaptation to both mega-environments would result in increased response to selection.

Introduction

Predictable genotype × environment (G×E) interactions, i.e., genotype × location (G×L) interactions across years (Allard and Bradshaw, 1964), within a target crop region can be exploited in plant breeding by grouping similar environments into subregions or mega-environments (CIMMYT, 1989) and targeting narrowly adapted cultivars to them (Cooper and DeLacy, 1994). A fair and straightforward comparison between this approach and selecting for broad adaptation to the undivided target region implies the absence of substantial differences in costs between the two strategies. With this limitation breeders face a dilemma. Division of the target region into more homogeneous subregions could increase genetic variance and the G to G×E interaction ratio within-subregion, but implies the division of testing resources, which could result in loss of precision in the estimation of genotype means within the smaller subregions (Atlin et al., 2000).

If G×L interaction explains a large portion of variability of the system under study, pattern analysis (clustering and ordination) can be used to group environments on the basis of similarity of cultivar relative performance (Cooper and DeLacy, 1994). The resulting environment grouping constitutes an objective basis for the division of the target region. However, pattern analysis models do not indicate if a greater response to selection will result from selection for specific adaptation to the identified subregions (Basford and Cooper, 1998). The theory of correlated response to selection (Falconer, 1989) can be used to determine the relative merits of selecting for specific versus broad adaptation (Atlin et al., 2000). If the objective is to improve performance in the subregion 1 (S1), selection may be undertaken directly in S1, or indirectly in the undivided target region (TR). The relative effectiveness of both strategies depends on the genetic correlation between genotype performance in S_1 and TR (*rg*) and the repeatability (*h*2) in each environment (Atlin et al., 2000). The predicted ratio of correlated response (*CR*) in S1 to indirect selection in TR relative to direct response (*DR*) in S1 may be used to determine if division of the target region is likely to increase response to selection and can be calculated as follows:

$$CR / DR = r_g \times \sqrt{h^2_{\mathrm{TR}} / h^2_{\mathrm{S1}}} \qquad [1]$$

If *CR* / *DR* is lower than 1, then division of the target region would result in an increase in the response to selection. The genetic correlation between the undivided region and its subregions can be estimated using the variance components derived from a genotype × subregion mixed model applied to a trial dataset (Atlin et al., 2000) using the formula:

$$r_g = \sqrt{\sigma^2_g / (\sigma^2_g + \sigma^2_{gs})} \qquad [2]$$

where $\sigma^2 g$ is the genetic variance component and $\sigma^2 gs$ the genotype × subregion interaction component of variance. Direct selection is likely to be effective when h^2S1 is high, although the testing resources were less than those of the undivided target region, and when there is substantial G×E interaction associated with the effect of the subregions on cultivar relative performance, causing *rg* to be low.

Pattern analysis of a reference set of 10 single-cross sunflower hybrids grown over 21 environments (three years) of Argentina (de la Vega et al., 2001) has revealed that northern (subtropical) and central (temperate) environments show repeatable differences in genotype discrimination, suggesting the possibility of exploiting specific adaptation to each region. In this study, an extended dataset consisting of 46 trials (seven years) was used to validate the subregions identified by de la Vega et al. (2001), and to determine if subdivision of the sunflower target region of Argentina in central and northern subregions is likely to increase response to selection.

Materials and Methods

Trial Data Set. A reference set of 10 sunflower single-cross hybrids was evaluated in 46 central, northern and managed environments of Argentina (1996/97 to 2002/03). Details of the experimental material, test environments, experimental design and measurements in a subset of these experiments are given in de la Vega et al. (2001). The locations (code, latitude) included in the trial dataset are 9 de Julio (9J, 35.5S), Charata (CH, 27.2S), Daireaux (DX, 36.6S), El Sombrerito (ES, 28.4S), Las Breñas (LB, 27.1S), Margarita (MA, 29.7S), Orán (OR, 23.0S), Paraná (PA, 31.7S), Reconquista (RE, 29.1S), Sampacho (SA, 33.3S), San Bernardo (SB, 27.3S), Sáenz Peña (SP, 26.8S), Villa Ocampo (VO, 28.5S), Venado Tuerto Oct planting (VT, 33.7S), Dec. planting (V2), and Dec planting with artificially extended photoperiod (VI). The hybrids composing the reference set were selected from the Advanta Semillas testing program based on their contrasting relative performance across environments for oil yield. This set includes former commercial hybrids widely planted in the central region (Contiflor 15, Contiflor 9 and TC 2001), commercial hybrids widely planted in the northern region (Morgan 734 and Aguará), and experimental hybrids that showed different patterns of adaptation to the mentioned commercial hybrids.

Ordination Analysis. Ordination analysis was applied to the genotype × trial array of best linear unbiased predictors for oil yield, following the procedures discussed by Cooper and DeLacy (1994). A biplot of the first two principal components was constructed from this analysis, to examine correlations among environments in the manner in which they influence the relative performance of the reference hybrids.

Correlated Response to Selection. A genotype × subregion analysis of variance was conducted for the entire dataset to examine partitions of the G and G×E interaction components of variance for oil yield. In this analysis, locations were classified as either northern-type or central-type, according to the results of pattern analysis. The phenotypic observation *yijkl* on hybrid *i* in replicate *l* of location *j*, year *k* and subregion *m* was modelled according to Atlin et al. (2000) as:

$$y_{imjkl} = \mu + s_m + (l/s)_{mj} + (y/s)_{mk} + (ly/s)_{mjk} + (r/sly)_{mjkl} + g_i + (gs)_{im}$$
$$[3] + (gl/s)_{mij} + (gy/s)_{mik} + (gly/s)_{mijk} + \varepsilon_{imjkl}$$

where μ is the grand mean; *sm* the effect of subregion *m*; (*l*/*s*)*mj* the effect of location *j* nested within subregion *m*; (*y*/*s*)*mk* the effect of year *k* nested within subregion *m*; (*ly*/*s*)*mjk* the effect of the interaction between location *j* and year *k* nested within subregion *m*; (*r*/*sly*)*mjkl* the effect of replicate *l* nested within the subregion-location-year combination *mjk*; g_i the genotypic effect of the hybrid *i*; (*gs*)*im* the interaction effect for hybrid *i* and subregion *m*; (*gl*/*s*)*mij* the interaction effect for hybrid *i* and location *j* nested within subregion *m*; (*gy*/*s*)*mik* the interaction effect between hybrid *i* and year *k* nested within subregion *m*; (*gly*/*s*)*mijk* the interaction effect between hybrid *i*, location *j* and year *k* nested within subregion *m*; and ε_i*mjkl* is the residual effect for hybrid *i* in replicate *l* of subregion-location-year combination *mjk* (experimental error). The effect of the subregion was considered fixed. All other terms are random. REML analysis was used to estimate variance components and their standard errors for the random terms in the model.

Genetic correlation between the target region and its constituent subregions was calculated according to Equation [2], using the variance components derived from mixed model [3]. The predicted ratio between *CR* and *DR* in central and northern subregions was calculated using Eq. [1]. Repeatabilities in the target region and each subregion were calculated according to Fehr (1987. p. 97), using the variance components derived from a conventional linear model for the analysis of genotypes, locations, years, and replicates, with all factors considered random (Atlin et al., 2000). To be consistent with the assumption of no increase of testing resources in the specific adaptation option, h^2SNorth and h^2SCentral were calculated using half of the total plot number (years × locations × replicates) used to calculate h^2TE.

Results

The 1st and the 2nd principal components of the biplot constructed from the ordination analysis accounted for 52% of the G×E interaction variation (Figure 1). The environment vectors covered the total range of Euclidean space, showing on average a strong negative association between central- and northern-type environments. The differences in genotype discrimination between central and northern environments were repeatable over years (Figure 1), confirming that both subregions are different mega-environments for the sunflower crop; the only exception being the central location Sampacho, which depending on the year associated with the northern- or central-type environments.

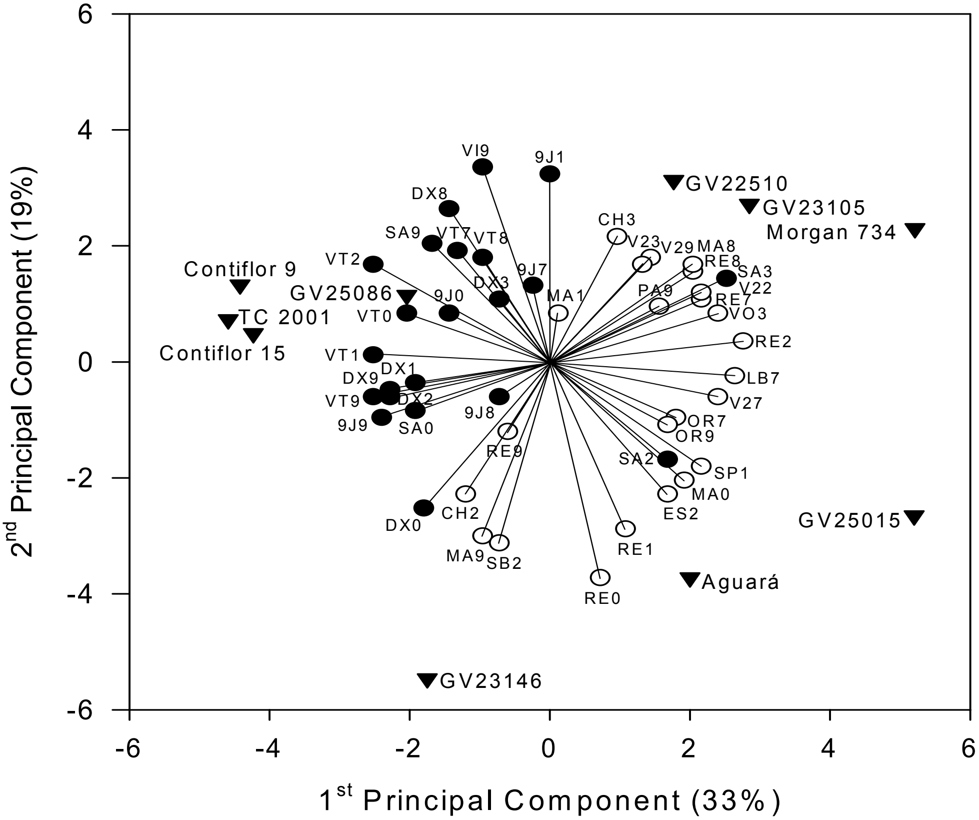

Figure 1. Biplot of the 1st and the 2nd principal components for oil yield of 10 sunflower hybrids grown in 46 trials. Genotypes are represented by points and environments by vectors. Black and open circles identify central- and northern-type environments, respectively. For any particular environment, genotypes can be compared by projecting a perpendicular from the genotype symbols to the environment vector. Acute angles between any two environment vectors indicate positive association, i.e., the environments compared influence the genotypic relative performance in a similar manner; 90º angles indicate no association; and angles greater than 90º indicate negative association (Kroonenberg, 1997). See Materials and Methods section for environment codes. Single years in trial codes (e.g., 7) refer to the summer season of that year, i.e., 1996/97 (from 1996/97 to 2002/03).

When normal October planting dates in the central location Venado Tuerto were delayed to December (trials V27, V29, V22, V23), they associated positively with the northern environments for oil yield. Planting date determines predictable G×E interactions, since this factor is under the control of the researcher or the farmer. In consequence, December plantings at Venado Tuerto were classified as northern-type environments for the analysis of correlated response to selection.

The variance component analysis showed a very high $\sigma^2 gs$ to $\sigma^2 g$ ratio in the undivided target region (Table 1). In consequence, an increase in σ^2_g and a decrease in total $\sigma^2 ge$ (i.e. G×L + G×Y + G×L×Y) to $\sigma^2 g$ ratio are observed when the target region is subdivided into central and northern subregions (Table 1). Calculated repeatabilities were 0.69 for the undivided target region (for a testing strategy of 14 locations, three years and three replicates) and 0.92 and 0.69 for the central and northern subregions, respectively (seven locations, three years and three replicates per subregion). The genetic correlation between the undivided target region and its constituent subregions was 0.36 and the calculated *CR*/*DR* ratios were 0.31 and 0.35 for the central and northern subregions respectively.

Table 1. Variance components (± standard errors) derived from the genotypes (G), subregions (S), locations (L), years (Y), and replicates (rep) model applied to the sunflower hybrid trial dataset.

Source of variation	Undivided target region		Within central region		Within northern region	
G	1394	± 4195	18142	± 9362	4491	± 2730
G×S	9645	± 5245				
G×L/S	2135	± 1272	2704	± 2024	238	± 1409
G×Y/S	1382	± 1168	0	± 1703	3106	± 1758
G×L×Y/S	12105	± 1952	12939	± 3234	12442	± 2498
rep/S×L×Y	2011	± 641	1668	± 1055	2286	± 793
Error	24202	± 1148	31656	± 2250	18197	± 1160

Discussion

Conversion of G×L interaction into $\sigma^2 g$ that can contribute to selection response is the main reason for the subdivision of large breeding targets (Comstock and Moll, 1963). If genotype × subregion interaction is large in comparison to the genotypic effect, then subdivision will be effective in increasing $\sigma^2 g$ in the subregions relative to the original undivided area (Atlin et al., 2000). Ordination analysis conducted on a dataset of 46 trials confirmed the subregion definition given by de la Vega et al. (2001). Subdivision of the sunflower target region of Argentina in northern and central subregions strongly increased the within-subregion $\sigma^2 g$ to total $\sigma^2 ge$ ratios (Table 1). The within-subregion $\sigma^2 gl/s$ to $\sigma^2 g$ ratios (Table 1) suggest that further subdivision is unlikely to result in increased response to selection. According to the calculated repeatabilities, the division of the testing resources did not result in loss of precision in the estimation of the within-subregion genotype means. This, together with a low *rg* between the target region and the subregions, resulted in low *CR* / *DR* ratios. These estimates are not strong evidence to assert that it is not possible to recombine some of the components of specific adaptation to the identified subregions in a way that will contribute to an improvement in broad adaptation, which is the main target of a breeder. However, they strongly support the notion that dividing the sunflower target region of Argentina into central and northern subregions and exploiting specific adaptation to both mega-environments would result in increased response to selection.

A point to note is the strong G×Y interaction observed in the northern region (3/4 of the G effect, Table 1). Northern-adapted hybrids can be classified as early-maturity (GV23105, GV22510, Morgan 734) or late-maturity hybrids (Aguará, GV25015, GV23146). Both groups showed contrasting relative performance in the direction of the 2nd PC of Figure 1, which discriminates between years in which the early hybrids show a better relative performance than the late ones, and years of opposite relative performance pattern. Analysis of private trial datasets (data not shown) suggests that the seasonal patterns of rainfall during grain filling, associated with the effect of the phases of El Niño Southern Oscillation, could be underlying this G×Y interaction effect. This hypothesis is under examination at present.

Acknowledgments

This research was supported by Advanta Semillas SAIC. I would like to thank Dr. Martín Grondona for his statistical advice and Dr. Antonio Hall for his comments on the manuscript.

References

Allard, R.W. and A.D. Bradshaw. 1964. Implications of genotype-environmental interactions in applied plant breeding. Crop Sci. 4:503-508.

Atlin, G.N., K.B. McRae and X. Lu. 2000. Genotype × region interaction for two-row barley yield in Canada. Crop Sci. 40:1-6.

Basford, K.E. and M. Cooper. 1998. Genotype × environment interactions and some considerations of their implications for wheat breeding in Australia. Aust. J. Agric. Res. 49:153-174.

CIMMYT. 1989. An account of how priorites are set among mega-environments from a breeding perspective. Internal Document Number 17. Mexico DF.

Comstock, R.E. and R.H. Moll. 1963. Genotype-environment interaction. *In*: Hanson, W.D., and Robinson, H.F. (eds.), Statistical Genetics and Plant Breeding. Publ. 982. National Academy of Sciences, National Research Council, Washington, DC. p. 164-194.

Cooper, M. and I.H. DeLacy. 1994. Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. Theor. Appl. Genet. 8:561-572.

de la Vega, A.J., S.C. Chapman and A.J. Hall. 2001. Genotype by environment interaction and indirect selection for yield in sunflower. I. Two-mode pattern analysis of oil and biomass yield across environments in Argentina. Field Crops Res. 72:17-38.

Falconer, D.S. 1989. Introduction to Quantitative Genetics, Third Ed. Longman Scientific & Technical, Burnt Mill, Harlow, England.

Fehr, W.R. 1987. Principles of Cultivar Development. Volume 1. Theory and Technique. MacMillan, New York.

Kroonenberg, P.M. 1997. Introduction to biplots for G×E tables. Research Report #51. Center for Statistics. The University of Queensland, Brisbane, Qld 4072 Australia. 20 p.