CBAB
CROP BREEDING AND
APPLIED BIOTECHNOLOGY

ARTICLES

Environmental stratification based on a 28 x 28 diallel of open-pollinated maize varieties

Cleso Antônio Patto Pacheco[1*], Heyder Diniz Silva[2], Manoel Xavier dos Santos[1], Paulo Evaristo de Oliveira Guimarães[1], Sidney Netto Parentoni[1], Elto Eugênio Gomes e Gama[1], Carlos Alberto Scapim[3], Walter Fernandes Meirelles[4], Hélio Wilson Lemos de Carvalho[5], and Pedro Abel Vieira Junior[6]

ABSTRACT – *The objective of this study was to assess the representativeness of the test environments used by the maize breeding program of Embrapa in the first phase of genotype evaluation. Ear weight of 378 hybrids from a diallel of 28 open-pollinated varieties (OPVs) evaluated in ten environments were used. The following environments were evaluated: two growing seasons (1991-92 and 1992-93), at three locations (Sete Lagoas, MG, Londrina, PR, and Goiania-GO); in two growing seasons (1991/92 and 1993/94) in Aracaju-SE; and in two growing seasons (1992-93 and 1993-94), in Ponta Grossa-PR. The complex part of the interaction accounted for nearly 75% of the genotype by environment interaction (G x E). The environments of Londrina-91/92, Ponta Grossa-93/94 and Aracaju-93/94 differed from the others and also from each other, as shown by stratification analysis. The phenotypic correlation between genotype means in the pairwise grouped environments, interpreted as coefficient of genotypic determination, indicated that non-genetic causes were responsible for 64.40% of the mean phenotypic variances. The results confirmed the discrimination of three major environmental groups, representing the Northeast (Aracaju), Central Southeast (Sete Lagoas, Goiania and Londrina) and South (Ponta Grossa) regions.*

Key words: Genotype by environment interaction, environmental stratification.

INTRODUCTION

A quantification of the percentage of the complex part of genotype by environment interaction (G x E) is very important to outline breeding strategies and to choose the test environments as well as the genotypes, according to the adaptability to a particular environmental condition (Vencovsky 1978).

When the complex part of the interaction is predominant, a genotype with superior performance in one environment will probably not perform as well in another, resulting in significant differences in yield, simply because of the genotype choice. The component of the G x E aside from the complex part is called the simple part of the interaction. When the simple part is predominant, the risk of selecting a wrong genotype is much lower because, even if the yield is lower, the ranking of cultivars would not change from one environment to another.

[1]Embrapa Milho e Sorgo, CP 151, 35701-970, Sete Lagoas, MG, Brazil. *E-mail: cleso@cpatc.embrapa.br
[2]Universidade Federal de Uberlândia, CP 593, 38400-902, Uberlândia, MG, Brazil
[3]Universidade Estadual de Maringá, 87045-000, Maringá, PR, Brazil
[4]Embrapa Soja, CP 231, 86001-970, Londrina, PR, Brazil
[5]Embrapa Tabuleiros Costeiros, CP 44, 49025-040, Aracaju, SE, Brazil
[6]Embrapa Negócios Tecnológicos, CP 151, 35701-970, Sete Lagoas, MG, Brazil

In most cases in the literature, the complex part of the interaction was the major component of the total interaction, and sometimes even confused with the G x E. Even in studies where the G x E caused some inconvenience in the selection but was not partitioned, the difficulty was due to the inconsistency of the cultivar response to environmental changes. The answer to this question would enable breeders to determine the environment where trials should be conducted in the most practical, inexpensive and efficient way within a subset of environments grouped by non-significant interaction, based on the similarity patterns of the genotype response and other aspects.

The objective of this study was to verify the representativeness of the main environments used in the maize breeding program of Embrapa Milho e Sorgo, for the first year of evaluation of hybrids

MATERIAL AND METHODS

Data from 441 treatments were used. The treatments comprised 28 open-pollinated populations (P), the 378 interpopulation hybrids (F_1) obtained from a diallel cross of these 28 populations, the first selfing generation (S_1) of each of the 28 populations, and 7 checks (see Pacheco et al. 2002a for a detailed description of treatments). Trials were conducted in experimental areas of the Brazilian Agricultural Research Corporation (EMBRAPA), in the growing seasons of 1991/92 and 1992/93, at the following EMBRAPA centers: Centro Nacional de Pesquisa de Milho e Sorgo (CNPMS) in Sete Lagoas (MG), Centro Nacional de Pesquisa de Soja (CNPSo) in Londrina (PR) and Centro Nacional de Pesquisa de Arroz e Feijão (CNPAF) in Goiania (GO), representing the Southeast (Londrina) and Mid-West (Sete Lagoas and Goiania) regions. Two trials were conducted at the Centro de Pesquisa Agropecuária dos Tabuleiros Costeiros (CPATC), in Aracaju (SE), representing the Northeast region, (growing seasons of 1991/92 and 1993/94) and two other trials were conducted at the headquarters of the Serviço de Negócios Tecnológicos (SNT – Ponta Grossa), in Ponta Grossa (PR), representing the Southern Region (growing seasons of 1992-93 and 1993-94), totaling 10 environments. The experimental conditions and data collection and analysis are described in detail by Pacheco et al. (2002b).

Environmental effects were considered random factors, since the geographical distribution of the experiments constituted a representative sample of the environmental conditions of maize-growing areas in Brazil. Population effects were considered fixed factors, since the populations represented a selected set of the best and/or most promising populations of the corn breeding program of the CNPMS, and are unlikely to be a representative random sample of the populations of the Maize Germplasm Bank (BAG).

Adjusted means of treatments involved in the diallel were used for the analysis. Joint analyses of environments, two at a time, were performed to estimate the complex part of interaction (C), using software Genes (Cruz 1997). The methodology was based on the expression of Cruz and Castoldi (1991):

$$C(\%) = \frac{\sqrt{(1-r_f)^3 Q_1 Q_2}}{Q_{12}} x100$$

where: C (%) is the percentage of the complex part of interaction; r_f is the phenotypic correlation between the means of the same genotype, in two environments; Q_1 and Q_2 are the mean squares of genotypes in environments 1 and 2 ; Q_{12} is the mean square of interaction between genotypes and environments, considering environments 1 and 2.

The stratification methodology of Lin (1982) was used, as proposed by Cruz and Regazzi (1994). It consists of the estimation of the sums of squares of the interaction between genotype and environments pairs, with subsequent grouping of the two environments with smaller and non-significant interaction, based on the F test. The process is then repeated, in an attempt to include a new environment in the first group of two environments, thus grouping the environments in groups of three , then four , and so on, until the F test is significant, indicating that no other environment can be included. The process should then be restarted with the still ungrouped environments.

RESULTS AND DISCUSSION

Despite the large number of treatments, the efficiency of the lattice compared to the randomized block design was low (Table 1). The performance of the lattice design was best in Aracaju - 91/92 (44.54%) and poorest in Ponta Grossa - 92/93 (0.08%). The use of a randomized block design would result, in the mean, in mean squares that would exceed the effective errors of lattice by 15.73%.

Table 1. Effective errors of the lattice design, values of the F test for the mean squares of treatments, efficiency of the lattice compared to the randomized block design, coefficient of variation and overall mean yield of ears, in the 10 evaluation environments

Environment	Location	Growing season	Effective error	F	Efficiency	C.V. (%)	Mean (kg ha⁻¹)
1	Sete Lagoas	91/92	787922.40	2.83 **	112.73	15.05 m	5902.90
2	Londrina	91/92	1713031.00	2.64 **	121.16	23.22 h	5635.94
3	Goiânia	91/92	970851.60	3.99 **	116.18	13.45 m	7327.36
4	Ponta Grossa	92/93	779618.50	7.42 **	100.08	9.07 l	9728.12
5	Aracaju	91/92	598172.80	2.74 **	144.54	21.57 m	3585.28
6	Sete Lagoas	92/93	807792.90	5.87 **	121.46	10.40 l	8640.34
7	Londrina	92/93	666805.20	5.57 **	103.72	11.53 m	7080.16
8	Goiânia	92/93	654484.60	3.49 **	102.65	16.29 m	4960.89
9	Ponta Grossa	93/94	1236070.00	5.08 **	112.72	14.57 m	7632.50
10	Aracaju	93/94	1693573.00	3.94 **	122.03	15.90 m	8185.71
Mean			990832.20		115.73	15.11	6867.91

** significant by the F test, at 1% probability
l, m and h represent the classification of C.V.(%) in low, medium and high, respectively, according to Scapim et al. (1995)

The coefficient of variation (CV) ranged from 9.07% to 23.22%, allowing the following classification of the experiments, according to Scapim et al. (1995): 2 environments - low; 7 environments - medium, and 1 environment - high. The mean CV of 15.11% is below the mean of 16.22% estimated by the authors for the trait ear yield, based on 66 other maize breeding trials.

The ratio of 2.86 times between the largest and the smallest effective error is well below the ratio of 7:1, indicated by Gomes (1990) as a threshold to perform combined ANOVA from trials with different residual mean squares.

The complex part of interaction (values above the diagonal in Table 2) accounted for a mean of about 75% of the G x E, indicating differences in the ranking of populations among environments. By the significance of the F test for the G x E in the environments, considered two by two, it was observed that the responses were always different (p<0.01) when the environments 2, 9 and 10 were involved (Londrina - 91/92, Ponta Grossa-93/94 and Aracaju-93/94).

Due to the imbalance of growing seasons, it was not possible to perform the analysis of variance, which would provide results on the genotype-year interaction,

Table 2. Part of the complex genotype x environment interaction, according to Cruz and Castoldi (1991), in % (in bold above the main diagonal) and estimates of simple correlation coefficients between genotype means, of the 10 pairwise environment combinations (in bold below the main diagonal)

	1	2	3	4	5	6	7	8	9	10	Mean
1 - Sete Lagoas (91/92)		84.1**	69.2	72.2**	83.2	**65.4***	78.6	83.4	75.2**	78.2**	76.6
2 - Londrina (91/92)	0.15		76.0**	81.2**	78.5**	86.3**	**84.3***	83.7**	87.7**	85.9**	83.1
3 - Goiânia (91/92)	0.47	0.40		69.5*	69.1	63.3	74.3	**76.6**	74.6**	80.7**	72.6
4 - Ponta Grossa (92/93)	0.38	0.34	0.51		68.9**	64.7	64.3	65.4	**69.3***	77.0**	70.3
5 - Aracaju (91/92)	0.30	0.17	0.42	0.36		70.6**	77.5	84.9	69.0**	**70.8***	74.7
6 - Sete Lagoas (92/93)	**0.46**	0.25	0.58	0.58	0.32		61.0	60.76	73.4**	79.1**	69.4
7 - Londrina (92/93)	0.34	**0.25**	0.45	0.57	0.31	0.60		71.4	75.8**	66.7**	72.7
8 - Goiânia (92/93)	0.30	0.18	**0.37**	0.49	0.26	0.53	0.46		65.2**	70.6**	73.6
9 - Ponta Grossa (93/94)	0.21	0.21	0.36	**0.48**	0.22	0.43	0.32	0.38		85.1**	75.0
10 - Aracaju (93/94)	0.16	0.24	0.27	0.37	**0.19**	0.35	0.45	0.30	0.28		77.1
Mean	0.31	0.24	0.43	0.45	0.28	0.46	0.42	0.36	0.32	0.29	

* and ** interaction between genotypes and the pairwise environment combinations, significant at 5% and 1% probability by the F test, respectively

considering all locations. However, the estimates based on data obtained in different years at the same location (values in bold in Table 2) show that G x E was non-significant in Goiania only, indicating the strong contribution of the effects of different growing seasons to the differential genotype response. It is possible that the genotype-year interaction at a same location is more important than the genotype-site interaction in a same year. Vencovsky and Torres (1988) found that these two forms of interaction were not correlated and may have distinct genetic bases.

It may seem strange that sometimes a lower value for the complex part of interaction was significant, e.g., 68.9% among environments 4 and 5, while a much higher one, e.g., 84.9% among environments 8 and 5, was not. It must be stressed, that although the data in Table 2 referred specifically to the percentage of the interaction between genotypes and environments, considered pairwise, the F test was based on the total magnitude of the G x E due to complex causes.

The agreement of estimates of simple correlation coefficients between genotype means (phenotypic correlation) in the pairwise combinations of the 10 environments (Table 2, below the main diagonal), with estimates of the respective percentages of the complex part of interaction was good. Comparing the means, on the sides of Table 2, it can be noted that the higher the correlation, the lower the contribution of the complex part, as expected.

The correlation coefficient between phenotypic means of genotypes (r_f) was estimated by the following expression:

$$r_f = \frac{Cov(F_1, F_2)}{\sqrt{V(F_1) \cdot V(F_2)}}$$ (i)

where: $Cov(F_1, F_2)$ is the covariance between means of the same genotype in environments 1 and 2; $V(F_1)$, $V(F_2)$ are the phenotypic variances of the genotype means within the environments 1 and 2, respectively.

If the environments were considered random, it can be assumed that $Cov(F_1, F_2) = \sigma^2_{g(12)}$, where $\sigma^2_{g(12)}$ is the component of genetic variance between genotype means in the mean of the two environments. Simultaneously, it can also be proved that

$$V(F_1) = \sigma^2_{F(1)} = \sigma^2_{g(1)} + (1/r)\sigma^2_{e(1)}$$

and

$$V(F_2) = \sigma^2_{F(2)} = \sigma^2_{g(2)} + (1/r)\sigma^2_{e(2)}$$

Based on a general expression proposed by Cruz and Regazzi (1994) considering only two environments, it can be shown that:

$$\sigma^2_{g(12)} + \sigma^2_{gxa(12)} = \left(\sigma^2_{g(1)} + \sigma^2_{g(2)}\right)/2$$

Based on this assumption, another, equally important expression can be inferred:

$$\sigma^2_{g(12)} + \sigma^2_{gxa(12)} + \overline{\sigma}^2_{e(12)} = \left(\sigma^2_{F(1)} + \sigma^2_{F(2)}\right)/2$$ (ii)

Thus, if the denominator of expression (i) is used to compute the phenotypic variances, it could be replaced by the mean values of the second expression (ii), and we would have:

$$r_f = \frac{\sigma^2_{g(12)}}{\sqrt{\left(\frac{\sigma^2_{F(1)} + \sigma^2_{F(2)}}{2}\right) \cdot \left(\frac{\sigma^2_{F(1)} + \sigma^2_{F(2)}}{2}\right)}}$$ (iii)

This new expression (iii) is still difficult to interpret, and could finally be re-written as follows:

$$r_f = \frac{\sigma^2_{g(12)}}{\sqrt{\left(\sigma^2_{g(12)} + \sigma^2_{ga(12)} + \overline{\sigma}^2_{e(12)}\right)^2}} = \frac{\sigma^2_{g(12)}}{\sigma^2_{g(12)} + \sigma^2_{ga(12)} + \overline{\sigma}^2_{e(12)}}$$ (iv)

In this final form (iv), it is easier to see that r_f, under certain assumptions, is an indirect measure of the heritability coefficient (h^2) and can therefore be interpreted as an indicator of the mean fraction of the phenotypic variance between two environments, which is due to genotypic causes, also means, between the two environments. Considering the genotype effects as fixed, the interpretation may still be true, although r_f would correspond to a genotypic coefficient of determination.

Thus, the highest phenotypic correlation of 0.60 between Londrina and Sete Lagoas in 1992-93, interpreted as genotypic coefficient of determination, indicates that 60% of the variation in the treatments occurred due to genotypic causes. In this case, one may say that the phenotypic value was a good predictor of the genotypic value. However, in the mean of the 10 pairwise environment combinations, it was observed that as a mean effect, non-genotypic causes were responsible for 64.40% of the phenotypic variation.

Data shown in Table 3 refer to the grouping of environments with non-significant G x E, according to Cruz and Regazzi (1994). The environmental stratification was in full agreement with the above explanations and conclusions on the complex part of G x E, evidencing that the results of the environments 2, 9 and 10 were different from the other seven and also from each other.

Table 3. Groups of environments with non-significant genotype-environment interaction for the treatments involved in the diallel

Groups	Environments [1]
I	1 5 8 3 7 6 4
II	10
III	2
IV	9

[1] where 1 and 6 correspond to two different growing seasons in Sete Lagoas (MG) and, respectively, 2 and 7 to Londrina (PR), 3 and 8 to Goiânia (GO), 4 and 9 to Ponta Grossa (PR) and 5 and 10 to Aracaju (SE)

The results of the clustering analysis represent exactly the order in which the environments were grouped. Thus, the experiments conducted in Sete Lagoas-91/92 and Aracaju-91/92 were those with the lowest G x E, and formed the first group. The two experiments conducted in Goiania were added to this first group followed by the environments represented by Londrina-92/93 and Sete Lagoas-92/93 and, finally by Ponta Grossa-92/93 (Table 3).

The three other environments were so different that they could not be grouped. The magnitudes of interaction were grouped in the following increasing order of divergence: Aracaju-93/94, Londrina-91/92 and Ponta Grossa 93/94.

It is noteworthy that these three isolated environments had the three highest effective errors in common, indicating that in these experiments the environmental influence was stronger than in the others (Table 1). In fact, the denominator of expression (iv) previously developed for the phenotypic correlation coefficient shows the importance of the average residual component in the environments involved. This average residual causes a decline in the phenotypic correlation among genotype means in both environments and increases the effects of the complex part of interaction.

This shows the trend of formation of three major groups, representing the regions Northeast, Central Southeast and South. Similar results would be expected in experiments conducted in Sete Lagoas, Londrina and Goiania, even if one of these three locations were eliminated. The Maize Breeding Program of the CNPMS could probably benefit if Sete Lagoas, Londrina or Goiania were replaced by another representative location of the South, North and Northeast, or any other location with a different adaptation condition from the three major groups mentioned above. Furthermore, the extent of the environmental influence represented by Ponta Grossa must also be evaluated, to adapt genotypes to conditions in the south of the country and of Aracaju with a view to the Northeastern region. The program should be focused on breeding for broad or specific genotype adaptation to these three major regions.

In a second analysis, these results ratify the choice of Sete Lagoas for the headquarters of the national corn breeding program, where the majority of the multidisciplinary team and research support is based, and where climate conditions allow crop cultivation in two growing seasons per year. More recently the selection and evaluation of specific lines at sites with diverging and controlled conditions, such as low and high Al^{+++} and / or phosphorus levels, have contributed to the development of more stable hybrids and synthetics, resulting from the crossing of lines that are efficient under stress conditions and responsive to improved cultivation conditions. In the coexisting activities of both cultivar establishment and evaluation at the CNPMS, the same human and financial resources are used. Due to the increasing economic difficulties, pressing on the public research institutions of agriculture and animal husbandry, the probable tendency is a shift towards prioritizing activities of cultivar establishment rather than of evaluation in the facilities in Sete Lagoas.

CONCLUSIONS

The complex part of the interaction accounted for about 75% of G x E. Differentiated responses were mainly caused by the environments in Londrina-91/92, Ponta Grossa-93/94 and Aracaju-93/94. The stratification ratified this information and showed that these three different environments differed from each other. The phenotypic correlation between genotype means in pairwise grouped environments, interpreted as coefficient of genotypic determination, indicated that non genetic causes were responsible for 64.40% of the mean phenotypic variation. The results confirm the discrimination of the environments in three major groups, representing the Northeast, Central Southeast and South. In view of the ease of setting up and conducting trials in the environment of Sete Lagoas and for being representative of the Mid-West and Southeast regions, this location is considered particularly advantageous, so that it is suggested to be given priority for the early phases of generating genotypes.

Estratificação de ambientes usando dados de um dialelo de 28 populações de milho

RESUMO – *O objetivo deste trabalho foi verificar a representatividade dos ambientes utilizados pelo Programa de melhoramento de milho da Embrapa na avaliação inicial de materiais. Foram utilizados dados de peso de espigas, de 378 híbridos de um dialelo de 28 populações, obtidos em 1991/92 e 1992/93, em Sete Lagoas-MG, Londrina-PR, e Goiânia-GO; em 1991/92 e 1993/94, em Aracaju-SE, e em 1992/93 e 1993/94, em Ponta Grossa-PR). A parte complexa representou cerca de 75% da interação genótipos x ambientes. Londrina-91/92, Ponta Grossa-93/94 e Aracaju-93/94 provocaram respostas diferenciadas, ratificadas pela estratificação, que ainda mostrou que esses três ambientes foram divergentes entre si. A correlação fenotípica entre médias de genótipos nos ambientes agrupados dois a dois, interpretada como coeficiente de determinação genotípico, indicou que causas não genotípicas foram responsáveis por 64,40% das variações fenotípicas médias. Os resultados confirmam a discriminação dos ambientes em três grandes grupos, representando as regiões Nordeste, Centro-Sudeste e Sul.*

Palavras-chave: interação genótipos x ambientes, estratificação de ambientes.

REFERENCES

Cruz CD (1997) **Programa Genes**. Editora UFV, Viçosa, 442p.

Cruz CD and Castoldi FL (1991) Decomposição da interação genótipos x ambientes em partes simples e complexa. **Ceres 38**: 422-430.

Cruz CD and Regazzi AJ (1994) **Modelos biométricos aplicados ao melhoramento genético**. Viçosa: UFV, Imprensa Universitária. 390 p.

Gomes FP (1990) **Curso de estatística experimental**. 13.ed. Piracicaba: ESALQ/USP. 468p.

Pacheco CAP, Santos MX, Cruz CD, Parentoni SN, Guimarães PEO, Gama EEG, Silva AE, Carvalho HWL and Vieira Jr PA (2002) Inbreeding depression of 28 maize elite open pollinated varieties. **Genetics and Molecular Biology 25**: 441-448.

Pacheco CAP, Santos MX, Cruz CD, Guimarães PEO, Parentoni SN and Gama EEG (2002) Topics on the genetics of maize yield stability. **Crop Breeding and Applied Biotechnology 2**: 345-354.

Scapim CA, Carvalho CG de and Cruz CD (1995) Uma proposta de classificação dos coeficientes de variação para a cultura do milho. **Pesquisa Agropecuária Brasileira 30**: 683-686.

Vencovsky R (1978) Herança quantitativa. In: Paterniani E (Ed.). **Melhoramento e produção do milho no Brasil**. Campinas: Fundação Cargill, Cap. 5, p. 122-201.

Vencovsky R and Torres RAA (1988) Estabilidade geográfica e temporal de alguns cultivares de milho. In: CONGRESSO NACIONAL DE MILHO E SORGO, 16º, 1986, Belo Horizonte, **Anais ....** Sete Lagoas: EMBRAPA/CNPMS. p.294-300.