Agronomía Costarricense 25(1): 21-32, 2001

GENOTYPE BY ENVIRONMENT INTERACTION AND GENETIC GAIN ON UNBALANCED *Pinus oocarpa* PROVENANCES TRIALS[1]

*Olman Murillo**

Key words: Genotype x Environment interaction, breeding, tree improvement, *Pinus oocarpa.*

RESUMEN

Interacción genotipo-ambiente y ganancia genética en ensayos de procedencia desbalanceados de *Pinus oocarpa.* Como parte del programa de mejoramiento genético de la Cooperativa para el desarrollo de los recursos genéticos de México y Centroamérica (CAMCORE), se estudiaron 9 procedencias de *Pinus oocarpa* provenientes de la zona seca de Guatemala y Honduras de 3 años de edad. Se utilizaron las mediciones de altura de los ensayos de procedencia/progenie en diseño de parcelas divididas, establecidos en 10 sitios a lo largo de Suramérica. No todas las procedencias ni tampoco las familias dentro de procedencias fueron plantadas en todos los sitios. Se observaron importantes cambios en el escalafón de las procedencias, y la magnitud e importancia de la interacción genotipo x ambiente se examinó a través del procedimiento de componentes principales. Los resultados se graficaron con el método de comparación de pares de Gabriel y se complementaron con los coeficientes de la correlación en el escalafón de Spearman. Se identificaron, con este procedimiento, 4 grupos de mejoramiento genético. Se analizaron las ganancias genéticas para cada caso así como los beneficios de agrupar los sitios en vez de mantener un único programa. El procedimiento de componentes principales mostró su efectividad para analizar correctamente la estructura correlacional de los datos y proveyó la información básica para separar los grupos de mejoramiento. Las interacciones GxE fueron significativamente reducidas y mayores ganancias genéticas fueron logradas dentro de cada uno de los 4 grupos creados, en comparación con el análisis de sitio general. Algunos efectos de confusión causados por el diferente número de familias dentro de procedencias a través de ambientes, pueden haber producido algún sesgo en los resultados.

ABSTRACT

Nine *Pinus oocarpa* provenances from the dry-zone of Guatemala and Honduras of 3 years of age, where compared by the CAMCORE's International Breeding Program. The data obtained was utilized for the development of an alternative GxE interaction analysis with unbalanced data. Height measurements from a provenance/progeny test on a split plot design at each of the 10 sites in South America were available. Neither all provenances nor all families within provenances were planted in all sites. Important rank changes occurred and the magnitude and importance of the GxE interactions were analyzed through the principal components procedure. Four major breeding groups were identified and other groups were also proposed on the basis of geographical criteria. Genetic gain estimations and benefits from grouping were analyzed by each possibility as well as against overall-site alternatives. The principal component analysis used proved to be a successful tool to estabilish correlational structure of data and provided the basic information for separating breeding groups. GxE interactions were significantly reduced and larger genetic gain estimations were achieved within each of the 4 groupings created, in comparison with overall site analysis. Some confounding effects caused by different family number within provenances across environments, may have produced some bias in the results.

1/ Recibido para publicación el 28 de marzo del 2001.

* Escuela de Ingeniería Forestal, Instituto Tecnológico de Costa Rica. Cartago, Costa Rica.
Correo electrónico: omurillo@itcr.ac.cr

lacional de los datos, así como para indicar la agrupación de sitios más apropiada. Con el empleo de estudios de esta metodología de análisis las interacciones genotipo x ambiente se redujeron sustancialmente y se obtuvo mucho mayores ganancias genéticas esperadas dentro de cada uno de los subgrupos de mejoramiento genético. El desbalance a nivel de familias dentro de las procedencias y a nivel de procedencias dentro de cada sitio, podría estar aún causando un sesgo en algunos de los estimados de los componentes de varianza.

INTRODUCTION

Genotype by environment (GxE) interactions can lead to the choice of a less than-desired seed source for planting at a particular site. The loss of gain depends on the differences between the chosen source and the "best" one (Matheson and Raymond 1984). As a result, interest in studies of genotype by environment interaction has increased in forest tree improvement programs (Gibson et al. 1983, Matheson and Raymond 1984, Lima 1987). Several different approaches have been developed to study GxE interaction since the early work of Yates and Cochran (1938) and Finlay and Wilkinson (1963). Methods of analyzing GxE interactions were reviewed by Freeman (1973), and new methods have been described since then by Westcott (1986) and Lin et al. (1986), among others. Matheson and Raymond (1984) presented a general review of this topic applied to tropical tree breeding programs. Several methodologies, mainly of the multivariate type, have been used to designate breeding regions and reduce the magnitude of the GxE interaction within regions (Abou-El-Fittouh et al. 1969, Burdon 1977). Principal component analysis is one multivariate approach that has been utilized in GxE interaction studies, and its statistical treatment can be found in literature (Rawlings 1988). Westcott (1986) cites works done with this type of analysis and mentions the difficulty interpreting the results where there was no obvious relationship to environmental conditions. Perkins (1972, 1974) and Perkins and Jinks (1968) developed an intensive study on GxE interaction methodologies based on inbred lines of *Nicotiana rustica*. In the forestry literature, Matheson (1976) has utilized principal component analysis with limited success because of the small number of environments analyzed. Barnes *et al.* (1984) compared different methodologies, including principal component analysis, working with *Pinus caribaea* provenances, and Kurinobu (1984) used this procedure for separating breeding regions for Japanese Larch wiht. However the proper statistical treatment of unbalanced data sets has not been always well documented. Nine *Pinus oocarpa* dry-zone provenances results at three years of age, from CAMCORE's International Breeding Program, were utilized for the development of an alternative GxE interaction analysis with unbalanced data. The information available for this tree species was mainly based on a few common provenances across a wide array of environments throughout the tropics (Dvorak 1986, Eguiluz 1986). With the exception of this study, the information reported has been based on provenance bulked seed collections with no family structure (mother-tree) within provenance. Among several other traits, total height is the less affected by stand density and is strongly correlated with volume, that makes it a variable usually utilized in breeding programs (Zobel and Talbert 1984).

MATERIAL AND METHODS

Data analyses were based on height measurements after 3 years in CAMCORE's (Central America and Mexico Coniferous Resources, headed by North Carolina State University) international project of *Pinus oocarpa*. The study consisted in comparing 9 provenances of *Pinus oocarpa* (from dry sites in Guatemala and Honduras) in which open-pollinated seed collections were made from natural populations. The provenances included in the study were Camotán, El Castaño, La Lagunilla, San Luis Jilotepeque all in Guatemala; La Campa, Las Crucitas, Guaimaca, San Marcos de Colón and Tablazón all in Honduras. A cheklot from Mountain Pine Ridge (Belize) was also included. Seedlings were planted at 10 very contrasting test sites in South America. The sites planted were Angatuba, Felixlandia and CPAC in research stations of EMBRAPA, Brazil; sites Jari 1, Jari 2 and Jari 3 all within lands of Jari Florestal company in Brazil; sites Aracruz1 and Aracruz3 within lands of Aracruz Florestal company in Brazil; one site in Popayán, Cali, owned by Cartón de Colombia company; one site in lands of Pizano/Monterrey company in Colombia. The experimental design was a split with 9 blocks. There were 6 to 8 open-pollinated provenances planted per test site, with an average of 8 families per provenance planted in 6 tree-row plots. Neither all of the same provenances nor all of the same families within provenances were present in all test sites. Data analyses were based on height measurements after 3 years in the field. Plot means were calculated for each family with more than 3 trees per plot. Least squares means were computed for each provenance using family plot means. No provenance with less than 5 families was included in the analyses. Ranks of provenances for each site were obtained by the least squares means procedures of the PROC GLM in the Statistical Analysis System (SAS 1985). Rank correlation for pairs of provenances among sites was performed using Spearman's procedure as a complement to principal component analysis.

$$R = \frac{\sum\left(R_i - \bar{R}\right)\left(S_i - \bar{R}\right)}{\sqrt{\sum\left(R_i - \bar{R}\right)^2 \sum\left(S_i - \bar{S}\right)^2}} \qquad (1)$$

Where:

R_i = the rank of the *i*th provenance in location 1.

$\bar{R}$ = the mean of the provenance's rank in location 1.

S_i = the rank of the *i*th provenance in location 2.

$\bar{S}$ = the mean of the provenance's rank in location 2.

All observations were standardized by the standard error at each site prior to the principal component analysis. This procedure removed the heterogeneity of regression common to the separate sites.

$$\text{Each value} = \frac{Y_{ijkl}}{\sqrt{\sigma^2}}$$

Where:

Y_{ijkl} = the *i*th tree of the *j*th family, within the *l*th provenance in the *k*th replication

σ^2 = error of the *j*th location.

Since all provenances were not planted at all sites, the balance of the data was improved by dropping from the analysis all provenances with less than 5 families. Also, the poorest represented provenances (Camotán and La Lagunilla) were excluded from the analysis, and the checklot (Mountain Pine Ridge, Belize) was included since it was present in 9 of the 10 sites. Those provenances at a given site, which had a high proportion of families poorly replicated across environments were also excluded from the analysis. Missing cells were estimated by the least squares procedure suggested by Steel and Torrie (1980).

The GxE interaction was then partitioned among cells (G_{ij}) as follows:

$$G_{ij} = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}$$

Where:

G_{ij} = GxE interaction accounted by the *i*th provenance at the *j*th site.
Y_{ij} = mean of the *i*th provenance at the *j*th site.
$\bar{Y}_{i.}$ = mean of the *i*th provenance across sites.
$\bar{Y}_{.j}$ = mean of the *j*th site across provenances.
$\bar{Y}_{..}$ = grand mean.

The GxE interaction data were subjected to principal components analysis. As a first step, the matrix data (standardized data) were subjected to a Singular Value Decomposition (SVD) procedure (SAS 1987), since the data corresponded to a nxp rectangular size matrix, with n=10 sites, and p=8 provenances. From the matrices produced by the former analysis on standardized values, an eigenanalysis was performed to get the respective eigenvalues and eigenvectors. With this information, the construction of the Gabriel's Biplot (Gabriel 1971, 1972 in Rawlings 1988) for sites for the first and second, and first and third principal components, displayed the correlational structure of the data. Representativeness of vector in the biplots were estimated as follows:

$$P = 100 \times \frac{\sqrt{(C_{i1})^2 + (C_{i2})^2}}{\sqrt{r_i}} \qquad (2)$$

Where:

P = proportion represented in the biplots.
C_{i1} = coordinate of the *i*th location on the first principal component.
C_{i2} = coordinate of the *i*th location on the second principal component.
r_i = correlation value of the *i*th location provided by the correlation matrix produced by the principal components analysis.

Results from Gabriel's Biplots, Spearman's rank correlation matrix and the rank of provenances by site, were analyzed to identify major breeding groups.

The procedure PROC GLM and type IV sums of squares (SAS 1985), were utilized in all combined site analyses of variance. Due to computational difficulties caused by the large amount of missing data (unbalanced data), the overall site analysis was split into separate analyses as follows: One overall analysis at the provenance level and 2 overall analyses at the family level without the provenance information. Total number of families was separated into 2 subsets, one constituted by the families from provenances 1 to 5, and the other composed of families from provenances 6 to 9. The 2 separate subsets were then analyzed and variance components and other estimates calculated separately. Both results were combined and averaged to obtain approximate overall family and family x site variance component estimates.

Similar combined analyses of variance were performed for each identified potential breeding group, with one analysis at the provenance level and one at the family level, without the provenance information. Variance components of interest were then estimated. Genetic gain was estimated overall and for each breeding group on an individual observation basis (Namkoong et al. 1966).

$$\text{Gain} = \frac{ik\sigma^2 A}{\sigma p} = \frac{i(0.5)(3 * \sigma^2 f)}{\sqrt{\sigma^2 + \sigma^2 rf(L) + \sigma^2 Lf + \sigma^2 f}} \qquad (3)$$

Where:

i = selection intensity. For the condition of this study it was assumed to be 1/1000, or *i*=3.367 (Becker 1984, Lindgren 1986).
k = fraction of the total additive genetic variance in the covariance of additive values. The fraction was assumed to be 0.5 since covariance (parent, offspring)= $\frac{1}{2}\ \sigma^{2a}$, with only the mother selected. This gain corresponds to selection of phenotypes in natural stands and utilization of their open-pollinated seeds for plantation establishment.
$\sigma^2 A$ = the additive variance, estimated by $3x\sigma^2$ fam variance component, since seeds collected

from mother trees could be a mixture of half-sibs and full-sibs.

σp = phenotypic standard deviation.

$\sigma^2 rf(L)$ = replication by family within location variance component.

$\sigma^2 t$ = within family plot variance component.

Genetic gain were similarly estimated on a family mean basis:

$$\text{Gain} = \frac{i(0.5)(3 * \sigma^2 f)}{\sqrt{\frac{\sigma^2 w}{nrL} + \frac{\sigma^2 rf(L)}{rL} + \frac{r\sigma^2 Lf}{L} + rL\sigma^2 f}} \quad (4)$$

Where the new term, $\sigma^2 w$, corresponds to the within plot mean variance component.

Genetic gain estimates by breeding group for selections at the family level were compared to the combined genetic gain over sites. Other possible grouping of sites such as geographical was also examined. Similar analyses and comparisons were performed among these groups.

Separate analyses of gain were conducted at the provenance level. The best 3 provenances per group assumed to be selected and the selection differential was calculated, as follows:

$$\text{Selection differential} = \bar{X}_{\text{best3}} - \bar{X}_{\text{group}}$$

Due to the higher risk of having biased results at the family level, effectiveness of grouping was measured at the provenance level by all possible pairwise comparisons, among the potential breeding groups and the combined data.

RESULTS AND DISCUSSION

The biplot with the first and second principal components accounted for 79% of the total variation. This was obtained by adding the first 2 values corresponding to the proportion explained by the first 2 eigenvalues given in Table 1. In the biplot each vector or line represents a single site (Figure 1). Each point or observation represents a single provenance. Numbers in parenthesis on

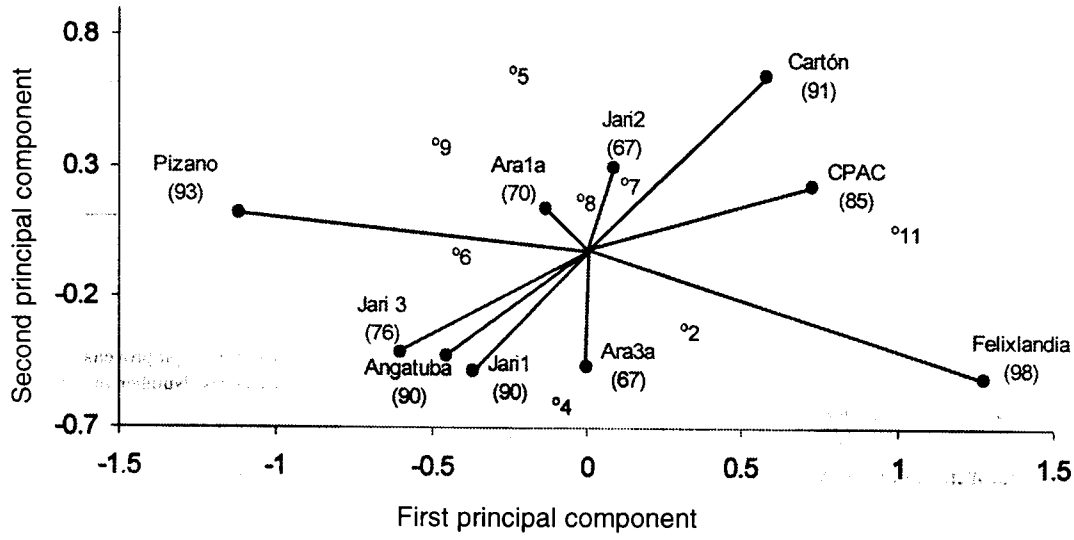

Fig. 1. First and second principal components for the GXE interaction data of CAMCORE's *Pinus oocarpa* provenances study. Vectors represent the different sites, and observations (circles) represent the provenances. Number in parenthesis indicates percentage of variation represented in this biplot for that vector.

Table 1. The column markers of eigenvalues and eigenvectors generated by principal components analysis on CAMCORE's *Pinus oocarpa* study for height at 3 years of age.

Eigenvalues	Variation % contributed by each eigenvalue	Site	Column markers*			Variation % contributed by each vector**
			D1	D2	D3	
4.47517	57.7500	Arala	-0.1326	0.1370	0.1669	92.83
1.65646	21.3700	Ara3a	0.0007	-0.4637	-0.4361	92.36
0.79004	10.1950	Cartón	0.5818	0.6468	0.0310	91.43
0.529954	6.8390	CPAC	0.7279	0.2284	0.2750	90.87
0.179532	2.3160	Felixlandia	1.2758	-0.5074	-0.1520	98.70
0.101588	1.3110	Angatuba	-0.4510	-0.4287	-0.1583	92.77
0.016608	0.2140	Jari 1	-0.3675	-0.4869	0.0245	90.43
0.000022	0.00028	Jari 2	0.0872	0.2997	-0.0814	69.46
8.0874 E-17	-------	Jari 3	-0.5993	-0.4115	0.6041	99.09
1.0497 E-17	-------	Pizano	-1.1208	0.1205	-0.2729	95.40

* Each one of the column markers represents the information accounted by the respective principal components or dimensions as follows: D1 as dimension 1, D2 as dimension 2, and D3 as dimension 3.

** Proportion of variation represented in total for each vector with the 2 biplots combined.

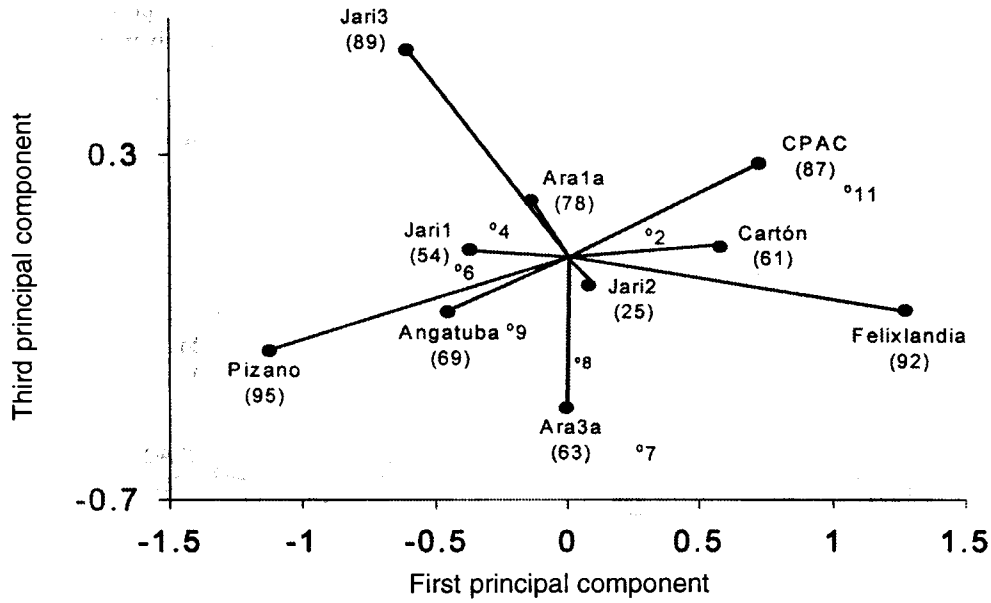

Fig. 2. First and third principal components for the GxE interaction data of CAMCORE's *Pinus oocarpa* provenances study. Vectors represent the different sites, and observations (circles) represent the provenances. Number in parenthesis indicates percentage of variation represented in this biplot for that vector.

each vector indicate how much of its variation was represented by this biplot. Including the third principal component the second biplot accounted for 89% of the total variation. The third principal component accounted for 10.19% of the total variation as shown by the third eigenvalue in Table 1. Since each principal component accounts for part of the total variation, some of the poorly represented vectors in the first biplot were better displayed in the second biplot (Figure 2). Therefore, a better interpretation of the correlational data structure was provided by the use

of information from both biplots. With the exception of vector Jari 2, all vectors had a very high proportion (>90%) of their respective variation represented, ensuring an appropriate interpretation of their correlational structure.

Most of the vectors exhibited more than 85% of their variation (indicated by the number in parenthesis on each vector) in the biplot with the first 2 principal components (Figure 1). Its length and departure represents the magnitude of the GxE interaction accounted for by each vector from the center of the biplot. Therefore, a small proportion of the total GxE interaction was accounted for Arala, Ara3a and Jari 2 vectors, as opposite to Felixlandia, Pizano, Cartón, Angatuba, CPAC and Jari 3.

Pizano, CPAC and Felixlandia vectors were clearly aligned with the axis of the first principal component (horizontal axis). Thus, primarily these 3 vectors defined the first principal component. Variation along the second principal component axis (vertical axis) was primarily due to sites Jari 1, Jari 2 and Ara3a. Site Cartón, Angatuba, Arala, and Jari 3 were located in between the 2 principal components and tended to be aligned mostly with the second principal component.

Similarities among sites (vectors) were determined by the magnitude of the angle formed between any pair of them, i.e., the smaller the angle between the vectors, the stronger is the positive correlation between the 2 sites. Highly positively correlated vectors indicated that provenances showed similar responses and rank positions in those environments. Conversely, non-correlated vectors (angles approaching 90°), or highly negatively correlated vectors (angles approaching 180°), indicate that provenances respond differently at sites.

The first 2 principal components showed vectors Cartón and CPAC well represented and highly positively correlated, as well as vectors Angatuba and Pizano (Figure 1). Vectors Jari 1 and Jari 3 seemed to be positively correlated; however, vector Jari 3 has only 76% of its original variation displayed. Therefore, vector Jari 3 may not be as highly correlated with vector Jari 1 as this biplot suggested. Vectors Jari 2 and Arala appeared to be positively correlated, but these 2 vectors had the poorest representation in this biplot; then, these 2 vectors may not be as well correlated as this biplot suggested. Vector Jari 2 appeared to be strongly and positively correlated to vector Cartón and well correlated to vector CPAC as well, but again, since vector Jari 2 was poorly represented in this biplot, its relationship must be interpreted with caution. Felixlandia vector was well represented in this biplot, it exhibited a strong negative correlation with vectors Angatuba and Pizano.

The provenances represented by dots in the biplots, tended to appear close to the region of the biplot where most of their GxE interaction was displayed. Thus, vectors that appeared in the opposite regions of the biplot, showed the opposite rank position for this provenance with respect to its rank position in their nearby vectors. The results may indicate that the farther the provenance was separated from the center of the biplot, the larger its GxE interaction was. Provenance 8 (San Marcos de Colón, Honduras) was found in previous studies to be one of the lowest exhibitors of GxE interaction and the only one that underwent no severe rank changes. In this biplot, it was the provenance plotted nearest the center. Since all other provenances had severe rank changes there was no other result that could support this finding. The ubication of provenance 4 (San Luis Jilotepeque, Guatemala) was near vectors Jari 1, Jari 3, and Ara3a (Figure 1). These were the sites where that provenance exhibited an upward rank change (Table 4). Vectors Cartón, CPAC, and Jari 2 appeared to be highly negatively correlated to this provenance, since their vectors on the biplot were in the opposite region. These sites were the ones in which provenance 4 ranked lowest (Table 4). Similarly, provenances 6 (Las Crucitas, Honduras) and 9 (Tablazón, Honduras) appeared near vectors Pizano, Angatuba, Arala, and Jari 3. In all these sites, these 2 provenances ranked their highest. Provenances 6 and 9, examined with respect to the position of vectors Felixlandia, Cartón and CPAC, showed that at these sites these 2 provenances ranked their lowest. This relationship was expected since these 2 groups of vectors were highly negatively correlated as shown by this same biplot.

Provenance 7 (Guaimaca, Honduras) seemed to have most of its GxE interaction manifested in vectors Ara1a, Jari 2, and Cartón; in its opposite region were vectors Jari 1 and Jari 3.

The same relationships found with the first 2 principal components were also observed, in general, in the biplot constructed on the first and third principal components (Figure 2). Some of the vectors such as Angatuba, Jari 1, Jari 2, and Cartón, had a poorer representation than in the first biplot (as reflected by the number in parenthesis in Figure 2). An important change showed vector Felixlandia, now close to vectors CPAC, Cartón and Jari 2. The biggest change occurred in vectors Jari 1 and Jari 3, with the later now well represented and distant from vector Ara3a. The provenances with important changes were only the ones that were highly related to the Jari 1 and Jari 3 vectors in the previous biplot. Therefore, provenances 7 (Guaimaca, Honduras), 8 (San Marcos de Colón, Honduras), and 9 (Tablazón, Honduras) underwent expected changes.

Some of the relationships between provenances and sites may not be adequately displayed by these 2 biplots because missing cells were estimated by least squares procedures in order to allow analysis of the 10 environments. Since these estimated values did not contribute to the GxE interaction, they did not figure in the principal components analysis. However, those sites with more estimated cells (as CPAC and Pizano) than the others, did not have a correspondingly poorer representation in either biplot. Then, the estimated cells did not produce an important negative impact on this principal components analysis of GxE interaction.

The total GxE interaction was still partially accounted for by heterogeneity of regressions among sites as well as the main portion, which was accounted for by true rank changes. Previous standardization of all single observations by their site standard error, removed only those within-site differences in response of the genotypes. Therefore, this proportion of the GxE interaction caused by heterogeneity of regression may have produced some difficulties in the interpretation of the data correlational structure, since it can not be explained only on a rank change basis.

Table 2. Speaman's rank correlations among sites with no estimated missing cells, of CAMCORE's *Pinus oocarpa* breeding program.

Site	Ara1a	Ara3a	Cartón	CPAC	Felix	Angatu	Jari 1	Jari 2	Jari 3	Pizano
Ara1a	1.000	0.100 5***	1.000** 5	0.800 5	0.100 5	0.700 5	-0.800 4	0.678 7	-0.100 5	-0.500 3
Ara3a		1.000	0.071 7	-0.200 4	0.321 7	0.821 7	0.200* 5	0.600 5	0.000 5	0.600 5
Cartón			1.000	0.800 4	0.357 7	0.250 7	-0.600 5	0.700 5	-0.700 5	-0.700 5
CPAC				1.000	0.800 4	0.200 4	-0.500 5	0.600 5	0.400 4	-1.000 2
Felix					1.000	0.071 7	0.200 5	-0.100 5	-0.700 5	0.100 5
Angatu						1.000	0.000 5	1.000 5	0.100*** 5	0.700 5
Jari 1							1.000	-0.800 4	-0.200 4	0.500 3
Jari 2								1.000	-0.100 5	0.500 3
Jari 3									1.000	0.400 4
Pizano										1.000

* Significant at P≤0.05, ** significant at P≤0.001.

*** Value represents total number of common observations utilized for calculating the correlation.

Table 3. Group composition proposed for CAMCORE's *Pinus oocarpa* breeding program.

Group I	Group II	Group III	Group IV
Pizano	Pizano	CPAC	Felixlandia
Jari 1	Angatuba	Cartón	CPAC
Jari 3	Jari 2	Jari 2	Ara3a
	Ara3a	Arala	

With the information from the 2 biplots the following breeding groups could be preliminarily combined:

I = Jari 1 and jari 3
II = Pizano, Angatuba, and tentatively Arala
III= Cartón, CPAC, and tentatively Jari 2
IV= Felixlandia, CPAC, and tentatively Cartón

Site Ara3a was not well represented by either biplot, therefore it had an uncertain group allocation. Similarly, sites Arala, Jari 1, and Jari 3, were also tentatively assigned to different groups. Some of the sites were included in more than one group according to their position in the biplots. Results from both biplots were compared with Speaman's rank correlation coefficients among sites (Table 2). These correlations were performed on the original data without estimating missing values.

High positive rank correlations existed among the same sites that were indicated by the biplots. The rank correlation between Jari 1 and Jari 3 was the only result that did not support the analysis based on biplots. This correlation was based on only 4 provenances common to both sites. If the same correlation was performed with the missing data estimated by least squares, as done in the principal components analysis, the result would be a rank correlation of 0.547, which would represent the relationship shown in the biplots. However, neither of the 2 vectors were well represented at the same time in the same biplot. Then, the relationship between these 2 sites must be interpreted with caution and with the use of both biplots.

Sites with a low contribution to the total GxE interaction had the poorest representation in the biplots. Therefore, their group allocation was uncertain. The use of the Spearman's rank correlation coefficients were valuable for assigning these sites. Sites Arala, Ara3a, Jari 1, Jari 2, and Jari 3, were finally assigned to groups based on their rank correlation among sites. Four groups were then identified (Table 3).

Combined analyses of variance within each group were performed (Table 4). All groupings reduced the GxE interaction variance components at both levels, provenance and family within provenance, by comparison with the

Table 4. Variance components for provenance, family, GxE interactions, and gain estimation for the best grouping alternatives suggested by the data analyses for CAMCORE's *Pinus oocarpa* breeding program.

Group	Variance components				Gain estimations			
	Provenance	LocXProv	Family	LocXFam	Gain[a]	Gain[b]	Ratio[c]	Ratio[d]
All sites	70.873	176.6706	476.678	248.513	18.248	74.048	2.490	0.520
I	76.373	150.9500	444.416	181.401	18.747	66.859	1.970	0.408
II	147.608	31.4106	546.134	105.761	27.948	89.872	0.212	0.193
III	225.602	-18.2260	539.332	193.038	25.897	83.467	0.000	0.357
IV	159.224	74.4305	452.771	271.442	25.64	74.481	0.467	0.600

a Gain estimate based on individual family observations.
b Gain estimate based on family means.
c Ratio obtained between (LocxProvenace)/Provenance variance components.
d Ratio obtained between (LocxFam)/Fam variance components.

Table 5. Effectiveness of the *Pinus oocarpa* (CAMCORE's breeding program) proposed breeding groups, measured by comparing selection differentials for the same top 3 provenances selected in different groups.

Group where the 3 best provenances were selected	Selection differentials for the 3 selected provenances in other groups				
	Overall	Group I	Group II	Group III	Group IV
Overall	10.805 (0.00)*	5.387 (-41.27)	14.425	10.368 (-4.62)	10.869 (20.32)
I	9.770		11.28	9.343 (-14.04)	
II	10.05	6.394 (-30.295)	15.613		
III	6.57 (-39.18)		12.725 (-18.49)		
IV			9.149		

* Number in parenthesis reflects the loss in percentage gain compared to the total possible gain achievable if selection were made in that group.

analysis over all sites. The provenance and family effects variance components were consequently relatively larger, and expected genetic gain was higher for each group compared with a single breeding program for all companies (sites).

Group I showed an important reduction in the GxE interactions, but expected gains were no greater than from maintaining all companies together in a single breeding program. Preliminary combined analyses (not shown above) between sites Jari 1 and Jari 3 alone, and in combination with Ara3a, produced poor genetic gain and reduced GxE interactions compared with the overall sites. Based on these results the best possible allocation was to combine sites Jari 1 and Jari 3 where at least the group was not inferior to the overall values. The lack of consistency in the data at the family level could have produced seriously biased results for these gain estimations because they may not be representative of the populations studied and the estimators based on a few degrees of freedom. In the light of this problem, other criteria based on values at the provenance level were utilized to show the benefits of grouping environments. At the provenance level, the imbalance of the data was not as serious as at the family level. Comparative analyses among groups at the provenance level would therefore be more reliable and representative for such data sets. One of criterion was the ratio between Loc x Provenance divided by Provenance (Table 4). Shelbourne (1972) suggested that ratios larger than 0.5 could be interpreted as having significant GxE interaction in terms of reducing genetic gain expectations in forest tree breeding. The application of this criterion produced, in almost all cases, ratios less than 0.5 at the provenance level. Group I was the only with a ratio of 1.97, but it was considerably smaller than that of overall sites.

Another criterion for measuring the benefits of grouping, was examination of the respective selection differentials from selecting the best 3 provenances within each grouping, and comparing these with the selection differentials derived from selecting the same three provenances at the other grouping. In Table 5 are shown the results of all possible comparisons among groups.

Comparisons should not be made among absolute values in Table 5, since they reflect the differences in average growth between sites within each group. Grouping produced an important positive increase in gain at the provenance level compared to selecting over all groups. This was specially clear in groups I and IV where the greatest differences occurred.

LITERATURE CITED

ABOU-EL-FITTOUH H.A., RAWLINGS J.O., MILLER P.A. 1969. Classification of environments to control genotype by environment interactions with an application to cotton. Crop Science 9(2):135-140.

BARNES R.D., BURLEY J., GIBSON G.L., GARCIA DE LEON J.P. 1984. Genotype-environment in tropical pines and their effects on the structure of breeding populations. Silvae Genética 336):186-198.

BECKER W.A. 1984. Manual of quantitative genetics. 4th edition. Academic Enterprises. Pullman, WA, USA. 176 p.

BURDON R.D. 1977. Genetic correlation as a concept for studying genotype-environment interaction in forest tree breeding. Silvae Genética 26(5-6):168-175.

DVORAK W.S. 1986. Provenance/progeny testing and breeding strategy for *Pinus tecunumanii*. *In:* Proc. IUFRO Conf. on Breeding theory, progeny testing and seed orchards. Williamsburg, VA., USA. p. 299-309.

EGUILUZ P.T. 1985. Taxonomic relationship of *Pinus tecunumanii* from Guatemala. Commonw. For. Rev. 65(4):303-313.

FINLAY K.W., WILKINSON G.N. 1963. The analysis of adaptation in a plant breeding programme. Australian Jour. Agric. Res. 14:742-754.

FREEMAN G.H. 1973. Statistical methods for the analysis of genotype-environment interactions. Heredity 31(3):339-354.

GIBSON G.L., BARNES R.D., BERRINGTON J. 1983. Provenance productivity in *Pinus caribaea* and its interaction with environment. Commonw. For. Rev. 62(2):93-106.

KURINOBU S. 1984. A methodological study on the analysis of progeny trial plantations of Japanese Larch. Bulletin of The Forest Tree Breeding Institute. N°2. Ibaraki, Japan. 60 p.

LIMA R. 1987. Stability of genotypes of *Pinus oocarpa* for specific gravity in three different environments in South America. MSc. Thesis, College of Forest Resources, North Carolina State University. Raleigh, North Carolina, USA. 48 p.

LIN C.S., BINNS M.R., LEFKOVITCH L.P. 1986. Stability analysis: where do we stand?. Crop Science 26(5):894-900.

LINDGREN D. 1986. Short Note: An approximate formula for selection intensity. Silvae Genética 35(5-6).

MATHESON A.C. 1976. Genotype x environment interaction in *Pinus radiata*. *In:* IUFRO Proc. Joint Meeting on Advanced Generation Breeding. Bordeaux France. 15 p.

MATHESON A.C., RAYMOND C.A. 1984. Provenance x environment interactions; its detection, practical importance and use with particular reference to tropical forestry. *In:* IUFRO Proc. of a joint workshop conference on provenance and genetic improvement strategies in tropical forest trees. Ed. by Barnes R.D. and Gibson G.L. Mutare, Zimbabwe. p. 81-117.

NAMKOONG G., SNYDER E.B., STONECYPHER R.W. 1966. Heritability and gain concepts for evaluating breeding systems such as seedling orchards. Silvae Genética 15:76-84.

PERKINS J.M. 1972. The principal component analysis of genotype-environmental interactions for multiple metrical traits. Heredity 29:51-70.

PERKINS J.M. 1974. Orthogonal and principal component analysis of genotype-environmental interactions for multiple metrical traits. Heredity 32(2):189-209.

PERKINS J.M., JINKS J.L. 1968. Environmental and genotype– environmental components of variability. III. Multiple lines and crosses. Heredity 23:339-356.

RAWLINGS J.O. 1988. Applied regression analysis: A research tool. Wadsworth & Brooks/Cole Advanced Books & Software. California, USA. 553 p.

SAS. 1985. SAS User's Guide: Statistics. 5th edition. Cary, North Carolina. USA. 956 p.

SAS. 1987. SAS User's IML. Cary, North Carolina, USA. 387 p.

SHELBOURNE C.J.A. 1972. Genotype –environment interactions: its study and its implications in forest tree improvement. *In:* IUFRO Proc. Genetics SABRAO Joint Symp. Govt. For. Exp. Stat. Tokyo, Japan. 27 p.

STEEL R.G.D., TORRIE J.H. 1980. Principles and procedures of statistics. 2d edition. McGraw Hill Book Company. 633 p.

WESTCOTT B. 1986. Some methods of analyzing genotype –environment interaction. Heredity 56:243-253.

YATES B.F., COCHRAN W.G. 1938. The analysis of groups of experiments. Jour. Agric. Sciences at Cambridge 28:556-580.