CHAPTER 12

Biometric procedures applied to genetic improvement of alfalfa

Cosme Damião Cruz
Leonardo Lopes Bhering
Reinaldo de Paula Ferreira

Introduction

In this chapter we address several biometric and statistic principles which allow the researcher to analyze experimental data and to generate useful information for an alfalfa breeding program. The use of these procedures in the several stages of a program of such nature will be illustrated and subdivided into three phases: start, in which the goal is to form a base population with broad genetic variability, with characteristics of agricultural interest and good adaptability; middle, in which there is concern about the conduction of segregating families which allow maximizing direct gains, indirect gains or simultaneous gains in important traits; and end, in which the improved genetic material is already available and the aim is to recommend it for broad regions or for specific regions, which makes studies on genotypes x environment interaction essential, as well as studies on adaptability and stability.

Base population formation

One of the main stages of the breeding program is the choice of parents which, after interbreeding, will form the base population in which the researcher will invest efforts searching for productive genetic material, with quality and good adaptation. Several criteria can be applied to choose the parents, especially the performance regarding characteristics of agricultural interest, combination capability and adaptability. Another key factor is the diversity among the group of parents, evaluated with the goal of identifying the hybrid combinations of greatest heterotic effect and greatest heterozygosity, so that in its segregating generations there is greater possibility of recovering superior genotypes. The formation of the base population is very important in the context of management and conservation of alfalfa germplasm, since it provides information on the available resources and helps the localization and interchange of such resources.

Genetic diversity has been evaluated through biometric techniques based on the quantification of heterosis, or through prediction processes. Among the methods based on biometric models, aiming at evaluating parental diversity, the diallel analyses are cited (Cruz, 2005).

Diallel analysis

Diallel analyses are designed for quantifying the genetic variability of the trait and for evaluating the genetic value of parents and the specific capability

and heterosis manifested in specific crossings. In diallels, it is necessary to evaluate the hybrid combinations between the parents. Diallel analysis has been routinely used in genetic improvement to evaluate a small number of parents (around ten). However, when a large number of potential parents is available to use in crossings to form a base population, obtaining experimental material can be impracticable and the study, impossible.

For diallels including only the hybrid combinations, the following statistic model has been adopted:

$$Y_{ij} = m + g_i + g_j + s_{ij} + \varepsilon_{ij},$$

where

Y_{ij} : mean value of hybrid ij (i, j = 1, 2, ... p, $i < j$)

m: general mean

g_i, g_j: effects of the general combination capability (GCC) in the i-th and j-th parents, respectively

s_{ij}: effect of the specific combination capability (SCC) for crossings between parents of orders i and j and

ε_{ij}: average experimental error

Considering, as illustration, an outline involving four parents, we then have the diallel scheme presented in Table 1.

Table 1. Diallel scheme including hybrids F1 of four parents.

Parent	1	2	3	4
1	-	Y_{12}	Y_{13}	Y_{14}
2		-	Y_{23}	Y_{24}
3			-	Y_{34}
4				-

The effects can be estimated through the following formulas:

$$\hat{m} = \frac{2}{p(p-1)} Y_{..}$$

$$\hat{g}_i = \frac{1}{(p-2)} [Y_{i.} - (p-1)\hat{m}] = \frac{1}{(p-2)} Y_{i.} - \frac{2}{p} Y_{..} = \frac{1}{p(p-2)} [pY_{i.} - 2Y_{..}]$$

$$\hat{s}_{ij} = Y_{ij} - (\hat{m} + \hat{g}_i + \hat{g}_j) = Y_{ij} - \frac{1}{(p-2)} (Y_{i.} + Y_{.j}) + \frac{2}{(p-1)(p-2)} Y_{..}$$

and the variance analysis is carried out as the scheme presented in Table 2.

Table 2. Variance analysis scheme for balanced diallels involving only F1 hybrids, according to the methodology proposed by Griffing (1956).

FV	GL	QM	F		E(QM)	
			Steady	Random	Steady	Random
GCC[1]	p-1	QMG	QMG/QMR	QMG/QMS	$\sigma^2_\varepsilon + (p-2)\phi_g$	$\sigma^2_\varepsilon + \sigma^2_s + (p-2)\sigma^2_s$
SCC[2]	p(p-3)/2	QMS	QMS/QMR	QMS/QMR	$\sigma^2_\varepsilon + \phi_s$	$\sigma^2_\varepsilon + \sigma^2_s$
Residue	f	QMR			σ^2_ε	σ^2_ε

[1] GCC = general combination capability. [2] SCC = specific combination capability.

The sum of squares (SS) will be given by

$$SS(GCC) = \sum \hat{g}_i Y_{i.} = \frac{1}{(p-2)} \sum Y^2_{i.} - \frac{4}{p(p-2)} Y^2_{..} \quad \text{and}$$

$$SS(SCC) = \sum_{i<j}\sum \hat{s}_{ij} Y_{ij} = \sum_{i<j}\sum Y^2_{ij} - \frac{1}{(p-2)} \sum_{i<j}\sum Y^2_{ij} (Y_{i.} + Y_{.j}) + \frac{2}{(p-1)(p-2)} Y^2_{..}$$

Based on this analysis, it is possible to evaluate the relative importance of additional genetic effects, expressed by the effects linked to GCC, as well as the ones due to dominance deviations, linked to SCC. This information is useful for establishing the best breeding strategy. When additional effects are marked, greater gains will be predicted, even when simpler breeding strategies are used. The results are also useful to point the parents with best performance and greatest genetic complementarity to be interbred.

Diallel crossings can also be used to obtain the heterotic potential of certain crossings. Madril et al. (2008) evaluated hybrids and parents of nine alfalfa germplasms which have been displaying importance in the formation of North American cultivars. After the hybrids were obtained, the existence of hybrid strength and great potential for continuous gains from recombining the elite materials used were verified.

Bolanõs-Aguilar et al. (2001) obtained a 7 x 7 diallel including the reciprocal in alfalfa to evaluate production of seeds and their components. In the analysis, the methodology described by Griffing (1956) was used. It was verified that GCC explained most of the evaluated traits, while SCC was only significant for the

trait “seed weight”. The large effect of the significance of GCC found by the authors suggests that gains can be obtained with successive breeding cycles. The absence of significance of the reciprocal for most characteristics indicated lack of maternal effect, except on the characteristic “seed production per plant”, in which the effect of the reciprocal was significant.

Genetic diversity

When a high number of parents are available, previous studies of predictive nature are recommendable, orienting the number and type of crossings in which to concentrate greater effort to obtain hybrids. Since they dispense the previous obtainment of hybrid combinations, predictive methods of diversity among parents have deserved considerable emphasis. Predictive methods are the ones based on differences – morphological, physiological, etc. – shown by parents for determining diversity, which is usually quantified through a dissimilarity measure (for instance, Euclidean distance and Mahalanobis distance). Inferring based on ecogeographical diversity is also an example of a predictive method for heterosis.

In the prediction of genetic diversity, several multivariate methods can be applied. Among them, we can quote principal components analysis and canonical variables analysis, and the agglomerative methods. The choice of the most adequate method has been determined based on the precision the researcher desires, on how easy the analysis is and on the way the data were obtained.

The methods based on principal components or in canonical variables allow studying the diversity in dispersion graphs, in which, usually, two cartesian axes are considered. In these studies, several characteristics are evaluated in a set of genotypes which, through statistical procedures, are summarized in few components (or canonical variables) and given by linear combinations of the original traits, independent of each other and with decreasing discrimination capacity, so that the first components (or canonical variables) explain the maximum of the variation existing in the original data.

Agglomerative methods differ from the others because they depend fundamentally on previously estimated dissimilarity measures, such as Euclidean distance or the generalized Mahalanobis distance, among others.

Grouping analysis

Grouping analysis aims at assorting, through some classification criterium, the parents (or any other type of sampling unit) in several groups, so that there is homogeneity within the group and heterogeneity between groups.

Alternatively, grouping analysis techniques have the goal of dividing an original group of observations into several groups, according to some similarity or dissimilarity criterium (Cruz et al., 2004).

In grouping analysis, several issues emerge. Thus, the final number of groups desired is questioned, as well as the adequacy of the participation achieved and the type of similarity measure to use. Regarding the number of groups desired, what is usually done is using several numbers of groups and, by some optimization criterium, selecting the most convenient one. To evaluate partition adequacy, it is common to use discriminant analysis and, regarding similarity measures, several are cited, but the most commonly employed in improvement are Euclidean and Mahalanobis distances for quantitative variables and Jaccard or Nei and Li indices for binary variables resulting from molecular markers studies.

The grouping process comprises basically two stages. The first one is related to estimating the similarity (or dissimilarity) measure between the parents and the second, to the adoption of the grouping technique for formation of the groups.

Dissimilarity measures

Genetic diversity studies aiming at identifying parents for hybridization have been carried out based on information on quantitative traits or on molecular markers.

In the case of quantitative traits, Average Euclidean distance ($d_{ii'}$) or generalized Mahalanobis distance ($D^2_{ii'}$) has been used to express genetic diversity. Although the latter is preferred, it can only be estimated when the residual covariances matrix is available, structured based on experimental assays with repetitions.

Generally, if X_{ij} is the observation in the i-th parent (i = 1, 2, ..., p), referring to the j-th trait (j = 1, 2, ..., n) studied, the Euclidean distance between two parents i and i' is defined through the expression

$$d_{ii'} = \sqrt{\sum_j (X_{ij} - X_{i'j})^2}$$

The generalized Mahalanobis distance is defined through

$$D^2_{ii'} = \delta' \Sigma^{-1} \delta,$$

where

δ: vector of deviations between the average values of parents in relation to the variances studies and

Σ: residual variances and covariances matrix, obtained from previous analyses, according to an appropriate statistical model.

To illustrate the example involving the evaluation of 20 cultivars will be considered regarding seven traits: height, dry matter production, dry matter percentage, crude protein, neutral detergent insoluble fiber, in vitro digestibility of dry matter and crude fiber, whose data are presented in Table 3.

Table 3. Evaluation of 20 alfalfa cultivars regarding seven phenotypic traits.

Genotype	Block	HEI[(1)]	DMP	DM	CP	NDF	IVD	CF
1	1	80.0	4,164.75	96.30	25.57	63.81	74.32	1.81
1	2	88.6	4,312.50	96.19	28.46	64.15	75.36	1.89
2	1	88.6	4,057.75	94.65	20.43	45.68	65.94	1.72
2	2	85.4	4,929.17	95.81	19.07	52.50	66.09	1.80
3	1	82.4	4,701.58	92.15	20.61	65.23	76.92	1.82
3	2	78.4	4,426.33	94.22	20.25	64.19	76.80	1.78
4	1	66.0	2,967.00	92.12	19.74	49.78	65.49	0.94
4	2	64.6	2,672.58	92.98	16.41	68.59	63.25	0.97
5	1	66.6	2,362.92	94.23	18.06	48.21	66.82	1.34
5	2	61.2	2,754.25	93.01	19.25	53.10	62.91	0.87
6	1	63.8	3,180.75	91.93	19.03	54.15	62.40	1.02
6	2	66.8	2,797.67	94.51	20.53	48.98	72.63	0.84
7	1	57.0	1,697.25	93.16	17.81	51.36	68.22	0.94
7	2	51.4	2,393.42	93.18	18.13	60.79	63.91	0.96
8	1	61.8	2,475.83	96.36	20.69	47.69	62.54	1.18
8	2	60.2	1,889.42	94.40	22.08	42.31	68.08	0.69
9	1	61.4	2,921.33	94.58	19.86	49.37	65.28	1.08
9	2	62.0	3,757.33	93.91	19.68	51.04	65.56	1.04
10	1	70.0	2,506.75	92.18	19.68	47.60	65.83	0.97
10	2	58.2	2,640.50	94.01	20.75	50.26	69.73	1.00
11	1	61.8	3,305.42	95.55	22.44	46.82	65.07	0.93

Continued...

Table 3. Continued.

Genotype	Block	HEI[(1)]	DMP	DM	CP	NDF	IVD	CF
11	2	60.4	3,625.92	92.51	18.59	63.14	28.92	1.03
12	1	54.0	2,679.75	91.50	21.79	45.99	65.83	0.69
12	2	59.2	3,021.58	93.21	19.16	63.78	65.37	1.23
13	1	59.4	2,253.42	96.14	18.15	47.20	60.52	1.29
13	2	64.8	3,225.17	94.42	20.37	50.72	68.64	0.87
14	1	68.6	3,245.33	95.05	20.48	50.13	62.36	1.82
14	2	63.4	2,907.92	92.99	16.74	50.56	68.57	1.26
15	1	63.6	3,622.42	94.87	18.21	53.00	61.74	0.98
15	2	60.2	3,189.08	91.52	19.51	65.62	62.94	1.05
16	1	64.2	2,567.33	93.94	19.58	43.23	65.75	1.03
16	2	60.4	2,664.50	93.63	18.78	49.47	62.80	1.50
17	1	66.6	2,690.75	92.8	17.37	53.02	64.52	1.22
17	2	67.2	2,880.50	95.03	18.60	50.26	62.36	1.02
18	1	67.2	1,520.83	91.67	19.50	48.92	60.54	0.85
18	2	62.8	1,830.50	94.16	19.72	46.58	60.48	0.96
19	1	67.2	2,751.08	96.16	15.15	48.77	60.99	1.31
19	2	63.4	2,942.50	92.64	14.09	59.25	66.55	1.12
20	1	44.8	1,986.00	90.22	12.37	50.12	62.06	1.09
20	2	48.8	1,610.67	90.72	13.69	43.26	70.26	0.69

[(1)] HEI = height; DMP = dry matter production; DM = dry matter percentage; CP = crude protein; NDF = neutral detergent insoluble fiber; IVD = in vitro digestion of dry matter; and CF = stem/leaf.

To study genetic diversity it is necessary, preliminarily, to obtain the estimates of the residual variances and covariances matrix, through variance analyses. Once we have the average values and the Σ matrix, we obtain the 20 x 20 dissimilarity matrix, which can be later subjected to the grouping analysis. This will allow inferring the similar groups and the dissimilar groups.

Grouping techniques

Since it is desirable to have information regarding each pair of parents in the grouping process, the number of estimates of dissimilarity measures

is rather large, and that makes it impracticable to recognize homogeneous groups through the simple visual examination of those estimates. To perform this task, grouping methods are used.

Among the grouping methods most commonly used in plant breeding, we can quote the hierarchical ones and the optimization ones. Their description is presented as follows.

a) Hierarchical methods

In hierarchical methods, parents are grouped through a process which is repeated in several levels, until the dendrogram or tree diagram is established. In this case, there is no concern for the optimal number of groups, since the greatest interest is in the "tree" and in the branches obtained. Delimitations can be established by visual examination of the dendrogram, in which high level change points are evaluated, usually taking them as delimiters of the number of parents for a given group.

Hierarchical methods are also divided into agglomeration methods and divisive methods. Among agglomerative methods, we can quote the single linkage method; the complete linkage method; the average linkage or unweighted pair group method with arithmetic mean (UPGMA), weighed or not; the centroid method, also weighed of not; and the one proposed by Ward (1963). Among the divisive methods, the Edwards and Cavalli-Sforza (1965) is the most commonly known.

For the example being considered, grouping was carried out through UPGMA, based on the Mahalanobis generalized distance, achieving the result shown in Figure 1.

Touil et al. (2008) used the hierarchical method to classify 29 alfalfa populations from the Mediterranean, to evaluate the genetic diversity among these populations, using ISSR (*inter simple sequence repeat*) molecular markers. To calculate the genetic diversity among the various populations, the authors used the index by Rogers and Tanimoto. After the grouping analyses, four groups were formed and related to the origin of the evaluated populations.

In another study aiming at evaluating genetic diversity among alfalfa populations, Segovia-Lerma et al. (2003), through the UPGMA grouping technique were able to separate 30 genotypes belonging to nine groups with well recognized germplasm based on their geographical origin, using 34 AFLP primers for it.

b) Optimization methods

In optimization methods, the set of parents is parted into non-empty and mutually exclusive subgroups through the maximization of some previously

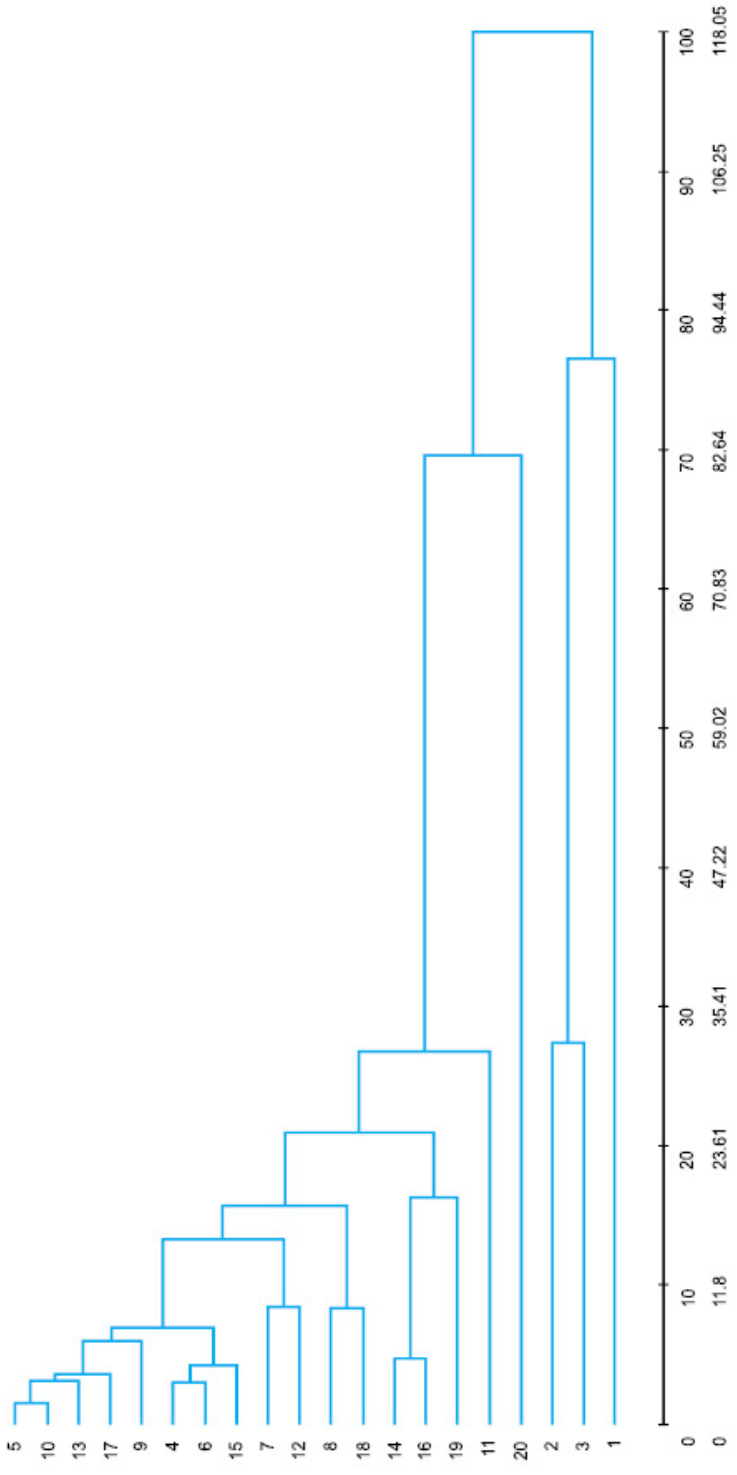

Figure 1. Dendrogram generated through the UPGMA grouping method, based on the Mahalanobis generalized distance.

set measure. One of the optimization methods most commonly employed in genetic improvement is the one proposed by Tocher, cited by Rao (1952).

In the Tocher method, the criterium adopted is that the average of the dissimilarity measures within each group must be smaller than the average distance between any groups. The method requires obtaining the dissimilarity matrix, upon which the most similar pair of parents is identified. These parents will form the initial group. From that, the possibility of including new parents is evaluated, adopting the criterium mentioned above.

For the example being considered, grouping was performed by the Tocher proposal, based on the generalized Mahalanobis distance, and the result obtained is indicated in Table 4.

We find the formation of four groups and that cultivars 1, 2, 3 and 20 are the most diverging ones in relation to the others. There is a group that differs in relation to these cultivars, but which still shows genetic diversity. The choice of cultivars to interbreed must consider the potential regarding the evaluated characteristics and the diversity. In this case, it is recommended to cross good cultivars belonging to different diversity groups.

Canonical variables analysis

This type of analysis requires more refined knowledge about multivariate statistical procedures, but it is easy to interpret and very useful in genetic diversity studies. It is based on generating new variables (named canonical

Table 4. Groups formed by Tocher methodology in an evaluation of 20 alfalfa cultivars.

Group	Cultivar
Ia	5 10 13 6 17 9 4 15 12 16
Ib	8 18
Ic	7
Id	11
Ie	14
IF	19
II	2 3
III	1
IV	20

variables), in which the information of the original measured variables is represented. These canonical variables are independent of each other and estimated so that the most variation is retained in descending order.

Thus, for the example being considered, the analyses allow concluding that with only two canonical variables (CV_1 and CV_2) it is possible to explain 81.2% of the variation found in the original data. These variables are defined through the following equations:

$$CV_1 = 0.177HEI + 0.001DMP + 0.039DM + 0.266CP + 0.063NDF + 0.029IVD + 0.981CF$$
$$CV_2 = 0.072HEI + 0.001DMP + 0.522DM - 0.855CP - 0.247NDF + 0.072IVD + 4.267CF.$$

The graphic analysis of the scores of the cultivars can be performed based on Figure 2. Again we note that cultivars 1, 2, 3 and 20 are the most diverging ones among the others.

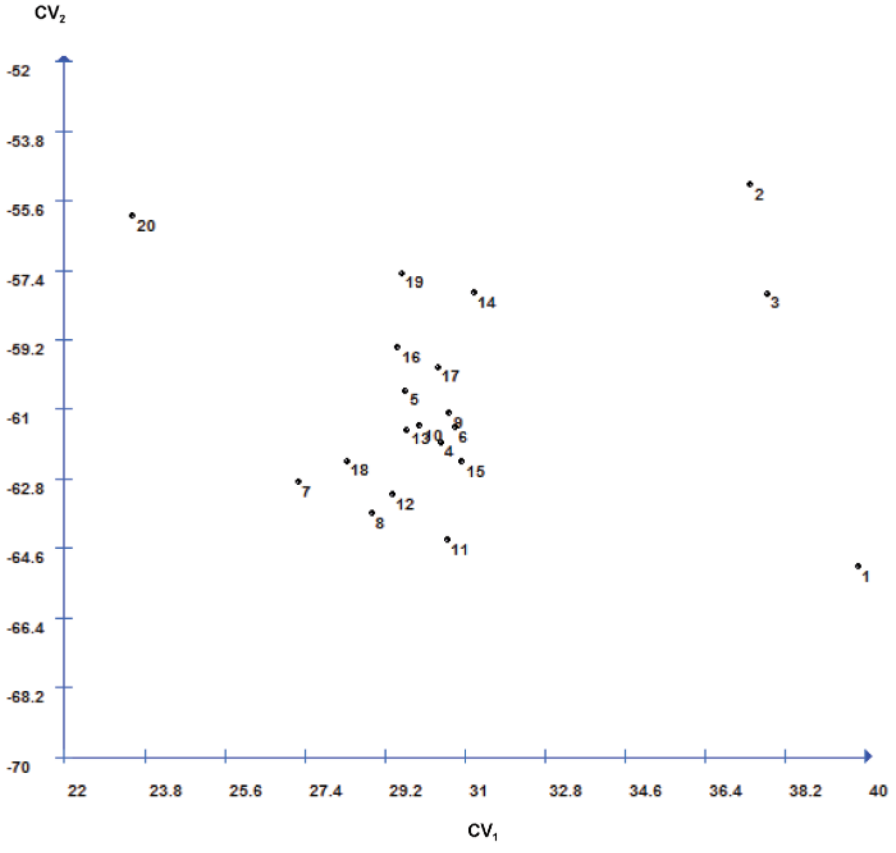

Figure 2. Graphic dispersion of scores of canonical variables obtained based on the linear combination of seven traits evaluated in alfalfa cultivars.

Conducting segregating populations

Predicting gains from selection

One of the great contributions of Quantitative Genetics is the evaluation of gains to be obtained through a given selection strategy. This information allows leading breeding programs, predicting their success, choosing or discarding populations and concentrating efforts in measuring traits of greater importance and greater gain potentiality. In this chapter, we address the prediction of gains obtained from selection in recurrent selection, in which selected individuals from an original population are tested and, themselves or others related to them, are recombined to obtain a new improved population in equilibrium.

To predict the gain from selection (*GS*), we use the expression

$$GS = p\hat{\sigma}_g\, hi,$$

where

p: parental control

$\hat{\sigma}_g$: genetic-additive standard deviation among test unities – normally corresponding to a fraction of the additive variance

h: heritability square root or accuracy of the selection process and

i: selection intensity

Some determinant factors of the gain from selection are: selection differential, selection intensity, parental control, genetic variability, environmental variance and genotypes x environments interaction.

Selection differential and selection intensity. One way to increase the gain from selection is to apply a higher intensity of selection, however, in very intense selection the population can present problems inherent to inbreeding, which is the consequence of crossing related individuals and is closely linked to the reduced size of samples.

Parental control. Parental control defines the similarity between test unities and the improved unities and, consequently, alters genetic covariance and the gain from selection.

Eberhart (1970) reported that parental control in a recurrent selection process can be defined by a function of the kinship relation between the selection unity used to identify superior individuals and the recombination unity used to obtain the improved population.

Genetic variability. Success of the breeding program depends on the existence of variability in the population. There is the concern to ensure wide variability in the work population, through the choice of diverging parents used in interbreedings to form the base population and through high specific capacity of combination.

Genetic variability is kept through adequate matings and proper samplings, so that the effective size of the population is not reduced.

Environmental variation. Phenotypic variation will be close to the genotypic variation when environmental variations are minimal. Thus, the environmental variation influences one of the main determining factors of the gain from selection, that is heritability. This coefficient is directly proportional to the additive genetic variability available in the population and inversely proportional to the phenotypic variation.

Genotypes x environments interaction. The existence of genotypes x environments interactions influences the gain to be achieved from selection. When the breeding program is restricted to a given environmental condition, this interaction is capitalized and, consequently, the fraction of used for predicting the gain is confused with the interaction σ2ga.

To illustrate, it will be considered that the 20 genotypes described in Table 3 constitute, instead of cultivars, half-sibling families derived from a breeding population. The gain achieved from selection of 30% of the best families can be estimated based on the information indicated in Table 5.

Based on the values of the means, of the average squares significance, of the variation coefficients and of the heritability (Table 5), the researcher can infer the genetic potential of the populations, the available variability, the experimental precision and the accuracy of the selective process. In this example, there is no possibility of gains in DM and IVD, since the genetic variability available through the half-sibling families is null. Increasing the genetic variability is recommended, by including new genotypes, by sampling another type of family or by better controlling the environmental influences, among other measures.

It is found that the experimental precision was adequate, that the greatest variation coefficient obtained reached 18% for CF and that there was significant genetic variance to be explored by selection in five out of the seven traits evaluated.

The averages of the evaluated families are presented in Table 6.

Table 5. Results of the variance analysis of seven agricultural traits evaluated in 20 half-sibling families of alfalfa.

FV	GL	QM						
		HEI[(1)]	DMP	DM	CP	NDF	IVD	CF
Block	1	19.04	197.86	0.16	0.17	195.67	0.09	0.05
Treatment	19	177.51**	1,280.33**	3.24^{ns}	14.41**	63.17*	64.44^{ns}	0.19**
Residue	19	11.51	109.93	2.01	1.98	30.31	45.13	0.04
Mean		65.06	2,953.24	93.71	19.26	52.71	65.11	1.16
VC (%)		5.21	11.23	1.51	7.31	10.44	10.32	18.0
h^2 (%)		93.51	91.41	37.96	86.23	52.00	29.96	77.7

[(1)] HEI = height, DMP = dry matter production, DM = dry matter percentage, CP = crude protein, NDF = neutral detergent insoluble fiber, IVD = *in vitro* digestibility of dry matter and CF = stem/leaf, VC = variation coefficient, h^2 = heritability.
ns,**,* Non-significant and significant, at 1% and 5% probability, by the F test, respectively.

Table 6. Average values of seven characteristics evaluated in 20 alfalfa half-sibling families.

Family	HEI[(1)]	DMP	DM	CP	NDF	IVD	CF
1	84.3*	4,238.63*	96.25	27.02*	63.98*	74.84	1.85*
2	87.0*	4,493.46*	95.23	19.75	49.09	66.02	1.76*
3	80.4*	4,563.96*	93.19	20.43*	64.71*	376.86	1.80*
4	65.3*	2,819.79	92.55	18.08	59.19*	64.37	0.96
5	63.9	2,558.59	93.62	18.66	50.66	64.87	1.11
6	65.3	2,989.21	93.22	19.78	51.57	67.52	0.93
7	54.2	2,045.34	93.17	17.97	56.08*	66.07	0.95
8	61.0	2,182.63	95.38	21.39*	45.00	65.31	0.94
9	61.7	3,339.33*	94.25	19.77	50.21	65.42	1.06
10	64.1	2,573.63	93.10	20.22*	48.93	67.78	0.99
11	61.1	3,465.67*	94.03	20.52*	54.98*	47.00	0.98
12	56.6	2,850.67	92.36	20.48*	54.89	65.6	0.96
13	62.1	2,739.30	95.28	19.26	48.96	64.58	1.08
14	66.0*	3,076.63	94.02	18.61	50.35	65.47	1.54*

Continued...

Table 6. Continued.

Family	HEI[1]	DMP	DM	CP	NDF	IVD	CF
15	61.9	3,405.75*	93.20	18.86	59.31*	62.34	1.02
16	62.3	2,615.92	93.79	19.18	46.35	64.28	1.27*
17	66.9*	2,785.63	93.92	17.99	51.64	63.44	1.12
18	65.0	1,675.67	92.92	19.61	47.75	60.51	0.91
19	65.3	2,846.79	94.40	14.62	54.01	63.77	1.22*
20	46.8	1,798.34	90.47	13.03	46.69	66.16	0.89
$\bar{X}_o$	65.06	2,953.24	93.71	19.26	52.71	65.11	1.16
$\bar{X}_s$[2]	74.98	3,917.79	-	21.67	59.70	-	1.57

[1] HEI = height; DMP = dry matter production; DM = dry matter percentage; CP = crude protein; NDF = neutral detergent insoluble fiber; IVD = in vitro digestibility of dry matter; and CF = stem/leaf.
[2] Average of the six superior families.

The estimates of the gain from selection, considering a breeding method in which the recombination only involves pollen from selected (parental control = 1), for the characteristic HEI, are the following:

$$\hat{\sigma}_g^2 = \frac{QMG - QMR}{r} = \frac{177.51 - 11.51}{2} = 83.00 \quad \text{and}$$

$$h^2 = \frac{\hat{\sigma}_g^2}{(QMG/r)} = \frac{83.00}{(177.51/2)} = 0.9351$$

For the selected percentage of 30%, we have $i = 1.159$ (Cruz, 2005). Hence, the gain from selection is estimated by

$$GS = i\, ph\, \hat{\sigma}_g = (1.159) \times 1 \times \sqrt{0.9351 \times 83.00} = 10.21$$

In percentage terms, we have

$$GS\ (\%) = \frac{GS \times 100}{average} = \frac{10.21 \times 100}{65.06} = 15.70\%$$

An alternative way to estimate the gain from selection is by the formula

$$GS = h^2\, DS = h^2\, (\bar{X}_s - \bar{X}_o) = 0.9351\ (74.98 - 65.06) = 9.28$$

In percentage terms, we have

$$GS\ (\%) = \frac{9.28 \times 100}{65.06} = 14.26\%$$

In Table 7, the estimates of genetic parameters are presented, as well as the predicted gain for variables HEI, DMP, CP, NDF and CF.

The gains from selection range between 6.9% and 29.9%. The two GS estimators, based on selection intensity or on the selection differential, provide similar estimates. It must be highlighted that the expression of GS based on the selection differential requires knowledge of the means of all the evaluated individuals, while the formula based on the knowledge of selection intensity does not require that knowledge; however, it must only be applied when the variable has normal distribution.

Table 7. Estimates of genetic parameters and predicted gain for the evaluated variables.

Trait[(1)]	X_o	X_s	h^2 (%)	GS = h^2DS	GS (%)	GS = $iph\hat{\sigma}_g$	GS (%)
HEI	65.06	74.98	93.52	9.28	14.26	10.21	15.70
DMP	2,953.24	3,917.79	91.41	881.73	29.86	847.69	28.70
CP	19.25	21.67	86.23	2.08	10.80	2.68	13.93
NDF	52.71	59.71	52.00	3.63	6.90	3.30	6.43
CF	1.16	1.57	77.74	0.32	27.14	0.28	24.3

[(1)] HEI = height; DMP = dry matter production; CP = crude protein; NDF = neutral detergent insoluble fiber; IVD = *in vitro* digestibility of dry matter; and CF = stem/leaf.

Direct and indirect selection of traits

As stated before, the success of a breeding program is primordially based on the existence of genetic variability, which enables selection and, consequently, the achievement of superior genetic materials to the breeder. The fast and efficient utilization of this variability is essential and studies about correlations constitute one of the paths to save time and to reduce efforts.

Estimating correlations is important to establish more adequate strategies for the condition of a breeding program and to evaluate the indirect responses in traits with low heritability or with problems of identification and measurement.

The correlations observed directly are phenotypic. It becomes necessary to distinguish their two causes: genetic and environmental. Genetic correlations are due mainly to pleiotropism and to the genetic changes in disequilibrium situations. Pleiotropism is the phenomenon through which a gene affects simultaneously two or more characteristics, so that if it is segregating it will cause concomitant variation in the characteristics involved. The correlation resulting from pleiotropism expresses the total effect of the segregating genes. Some pleiotropic effects can increase the characteristics, while others can reduce them. In other cases, the effects can increase some characteristics and reduce others, so that pleiotropism does not necessarily cause a correlation that can be detected.

Genetic linkage disequilibrium is the temporary cause of correlation and this can be altered in advanced generations due to breaks in the gene pools resulting from crossovers.

Phenotypic correlations are the ones obtained based on the means of the evaluated traits. When we consider the evaluation of two characteristics, X and Y, in *g* genotypes evaluated in *b* random blocks, we have the following statistical model:

$$X_{ij} \text{ or } Y_{ij} = \mu + g_i + b_j + \varepsilon_{ij}$$

The phenotypic correlation is obtained as follows:

$$r_f = \frac{Cov(\overline{X}, \overline{Y})}{\sqrt{V(\overline{X})\ V(\overline{Y})}}$$

where *X* and *Y* are variables that express the means of the genotypes in relation to traits X and Y, respectively.

To estimate the environmental correlations and the genotypic correlations, we must perform variance analyses and obtain the values of the mean squares (or variances). The mean products (covariances) are calculated using estimates of the mean squares obtained by the variance analysis of the sum of the values of *X* and *Y*, given by

$$Z_{ij} = X_{ij} + Y_{ij}$$

The scheme of variance analyses of variables *X*, *Y* and *X* + *Y* with the mean squares, indispensable for calculating the correlations, is presented in Table 8.

The mean products are obtained considering that

$$V(X + Y) = V(X) + V(Y) + 2\ Cov(X + Y),$$

Table 8. Scheme of variance analyses of traits X, Y and of the sum X + Y, evaluated in random blocks involving g genotypes.

FV	GL	X		Y		X + Y	
		QM	E(QM)	QM	E(QM)	QM	E(QM)
Blocks	r – 1						
Genotypes	G – 1	QMGx	$\sigma_x^2 + r\sigma_{gx}^2$	QMGy	$\sigma_y^2 + r\sigma_{gy}^2$	QMGx+y	$\sigma_{x+y}^2 + \sigma_{g(x+y)}^2$
Residue	(r-1)(g-1)	QMRx	σ_x^2	QMRy	σ_y^2	QMRx+y	σ_{x+y}^2

so that

$$PMG_{x,y} = \frac{QMG_{(x+y)} - QMG_x - QMG_y}{2} \quad \text{and}$$

$$PMR_{x,y} = \frac{QMR_{(x+y)} - QMR_x - QMR_y}{2}$$

The correlations can then be obtained as follows:

Environmental correlation:

$$r_a = \frac{PMR_{xy}}{\sqrt{QMR_x\ QMR_y}}$$

Genotypic correlation:

$$r_g = \frac{\hat{\sigma}_{g(x,y)}}{\sqrt{\hat{\sigma}_{gx}^2\ \hat{\sigma}_{gy}^2}}$$

where

$$\hat{\sigma}_{g(x,y)} = \frac{PMG_{x,y} - PMR_{x,y}}{r}$$

$$\hat{\sigma}_{gx}^2 = \frac{QMG_x - QMR_x}{r}$$

$$\hat{\sigma}_{gy}^2 = \frac{QMG_y - QMR_y}{r}$$

Phenotypic correlation:

Phenotypic correlation can also be obtained based on the mean squares, given by

$$r_f = \frac{PMG_{x,y}}{\sqrt{QMG_x\ QMG_y}}$$

For the example being considered, the correlation estimates obtained are shown in Table 9.

Table 9. Estimates of phenotypic and genotypic (in brackets) correlations, above the diagonal, and environmental, bellow the diagonal, between the combinations of seven characteristics evaluated in alfalfa.

Trait	HEI[(1)]	DMP	DM	CP	NDF	IVD	CF
HEI	1	0.807** (0.868)	0.608** -	0.576** (0.616)	0.416 (0.623)	0.424 -	0.852** (0.996)
DMP	0.059	1	0.463* -	0.493* (0.554)	0.601** (0.801)	0.302 -	0.796** (0.909)
DM	0.058	0.016	1	0.617** -	0.056 -	0.080 -	0.509* -
CP	0.241	0.013	0.139	1	0.359 (0.518)	0.251 -	0.427 (0.580)
NDF	-0.104	0.241	-0.318	-0.606	1	0.323 -	0.423 (0.390)
IVD	0.125	-0.223	0.231	0.480	-0.516	1	0.536* -
CF	0.023	0.215	0.301	-0.276	0.535	-0.350	1

(1) HEI = height; DMP = dry matter production; DM = dry matter percentage; CP = crude protein; NDF = neutral detergent insoluble fiber; IVD = in vitro digestibility of dry matter; and CF = stem/leaf.

**, * Significant, at 1% and 5% probability, by test t, respectively.

Correlated response to selection

The existence of genetic correlation between traits means that selection in one characteristic causes changes in others. Evaluating the direction and

dimension of these changes is essential to have, by the end of a breeding program, genetic materials with superior behavior in a number of characteristics.

Response expected from a trait Y, when selection is applied to a trait X, can be estimated by the following expression:

$$RY(X) = \hat{\beta}_g \, RX = \hat{\beta}_g \, \hat{\beta} \, DS_X$$

where

RX : direct response in trait X, given by $\hat{\beta} DS$;

$\hat{\beta}_g$: regression coefficient that measures variation in genetic values of the trait Y, calculated by

$$\hat{\beta}_g = \frac{Cov(X, Y)}{\hat{\sigma}^2_{gx}} = r_g \frac{\hat{\sigma}_{gy}}{\hat{\sigma}_{gx}}$$

$\hat{\beta}$: regression coefficient that measures variation in the genetic values in the improved population, regarding trait X, with changes caused by phenotypic selection in test unities, calculated by

$$\hat{\beta} = \frac{Cov_g(UM_X \, , \, UT_Y)}{\sigma^2_{fx}}$$

UM and *UT* are unities or individuals from the improved and from the test populations, respectively.

Regarding trait Y, we can estimate the direct gains through the formula

$$RY = p \; i_y \; h_y \; \hat{\sigma}_{gy}$$

and the indirect gains through the formula

$$RY(X) = p \; i_x \; h_x \; \hat{\sigma}_{gy} \; r_{gxy}$$

Therefore, the efficiency of indirect selection in relation to direct selection is given by the following ratio:

$$\frac{RY(X)}{RY} = \frac{p \; i_x \; h_x \; \hat{\sigma}_{gy} \; r_{gxy}}{p \; i_y \; h_y \; \hat{\sigma}_{gy}}$$

With the same selection intensity in traits X and Y, we have

$$\frac{RY(X)}{RY} = \frac{h_x \; r_{gxy}}{h_y}$$

from which we conclude that $RY(X) > RY$ if $r_g h_x > h_y$. Thus, response from indirect selection will be compensatory when the main trait (Y) has low heritability and when an auxiliary trait (X) of easy measurement, high heritability and high correlation to the main trait is available.

Based on the data presented in Table 5, we can obtain the correlated responses shown bellow.

Correlated response in Y, by selection in X

Let's consider the selection of 30% of the studied genotypes. Variable HEI will be called X and variable DMP, Y. Based on the results of the variance analysis presented in Table 5, we have

$$\hat{\sigma}_{gy} = \sqrt{585{,}198.8485} = 764.983;$$

$$h_x^2 = \frac{\hat{\sigma}_{gx}^2}{(QMT_X)/r} = \frac{83.00}{(177.5082/2)} = 0.9351 \text{ then } h_x = 0.9670 \text{ and } r_{gxy} = 0.8685$$

For the selected percentage of 30%, the selection intensity (i) value equals 1.159.

Thus, $RY(X) = 1.159 \times 0.9670 \times 764.983 \times (0.8685) = 744.6145$ assuming $p = 1$.

The direct response in Y is given by

$$RY = i_y h_y \hat{\sigma}_{gy}$$

where:

$$h_y^2 = \frac{\hat{\sigma}_{gy}^2}{(QMT_Y)/r} = \frac{585{,}198.8485}{1{,}280{,}331.3806/2} = 0.9141 \text{ then } h_y = 0.9561$$

thus $RY = 1.159 \times 0.9561 \times 764.983 = 847.6929$.

An alternative way to obtain the direct and correlated response in trait Y is by using the expressions based on selection differentials, obtained based on the means presented in Table 6.

Direct response is given by

$$RY = h_Y^2 DS_Y = h_Y^2 (\bar{Y}_S - \bar{Y}_O) = 0.9141 \ (3{,}917.7983 - 2{,}953.2437) = 881.7346$$

Indirect response is given by

$$RY(X) = h_Y^2 DS_{Y(X)} = h_Y^2 (\bar{Y}_{(X)} - \bar{Y}_O), \quad \text{in which}$$

$\overline{Y}_{(X)}$ is the mean of genotypes for trait Y, whose superiority was identified through the good performance in trait X. For the example at issue, $\overline{Y}_{(X)}$ is the mean of progenies 1, 2, 3, 4, 14 and 17.

Hence, we have

$$\overline{Y}_{(X)} = \frac{4{,}238.63 + 4{,}493.46 + 4{,}563.96 + 2{,}819.79 + 3{,}076.63 + 2{,}785.63}{6} =$$

$$= \frac{21{,}978.1}{6} = 3{,}663.01$$

and thus

$$RY(X) = 0.9141 \times (3663.013 - 2953.2437) = 648{,}8.$$

In some situations, the estimates of correlated responses obtained by the two estimators mentioned present results showing a great deal of discrepancy, disagreeing in dimension and, most surprisingly, in sign. Since the expression based on selection differentials is supported by just the ratio between two traits of the selected genotypes, it seems to be a good option used in breeding studies. In this example, we find that

$$RX(Y) = i_y\, h_y\, \hat{\sigma}_{gx}\, r_{gxy} = 1.159 \times 0.9561 \times \sqrt{83.00} \times (0.8685) = 8.768$$

and

$$RX(Y) = h^2_X\, DS_{X(Y)} = 7.1760.$$

When selecting trait Y, genotypes 1, 2, 3, 9, 11 and 15 were identified as superior. Among these genotypes, only 9, 11 and 15 produce less than the mean of trait X, while the others produce more, and genotypes 1, 2 and 3 were also selected for good performance in both Y and X. Given these facts, we can expect that the correlated response provides substantial gains in both characteristics evaluated.

Recommendation of cultivars

In a given environment, the phenotypic manifestation is the result of the action of the genotype under influence of the environment. However, when a number of environments are considered, we detect, in addition to genetic and environmental effects, an additional effect, given by the interaction between these effects.

Evaluating the interaction of genotypes x environments becomes very important to breeding because, in case it exists, there is the possibility for the

best genotype in one environment not to be so in another. This fact influences gain from selection and makes it difficult to recommend cultivars with broad adaptability. Because of the importance of this interaction, it is the breeder's job to assess its dimension and significance, quantify its effects in breeding techniques and the technology diffusion strategies, and to provide subsidies which allow adopting procedures for to minimize or utilize it.

Even though studies on interactions of genotypes x environments have great importance for breeding, they do not provide detailed information about the behavior of each genotype faced with environmental variations. For this purpose, adaptability and stability analyses are carried out, through which it becomes possible to identify cultivars with predictable behavior and that are responsive to environmental variations, in broad or specific conditions.

There are currently over ten methodologies of adaptability and stability analysis aiming at evaluating groups of genotypic materials tested in a number of environments. These methodologies are based on the existence of interactions and differ from the stability concepts adopted and from certain statistical principles employed. The choice of a method of analysis depends on the experimental data, mainly the ones related to the number of available environments, on the precision required and on the type of information desired. It should also be understood that some methods are alternative, while others are complementary and can be used together.

Eberhart and Russell (1966) suggested performing the adaptability and stability analysis based on a simple linear regression model. By these authors' proposal, the following equation is adopted:

$$Y_{ij} = \beta_{0i} + \beta_{1i} I_j + \delta_{ij} + \bar{\varepsilon}_{ij}$$

where

Y_{ij}: mean of genotype i in environment j

β_{0i}: general mean of genotype i

β_{1i}: linear regression coefficient, that measures the response of the i-th genotype to environmental variation

I_j: codified environmental index $\left(\sum_j I_j = 0\right)$

δ_{ij}: regression deviation and

$\bar{\varepsilon}_{ij}$: average experimental error

By this methodology, both the regression coefficients of phenotypic values for each genotype in relation to the environmental index, and the deviations

of such regression, would provide estimates of stability and adaptability parameters. The concepts involved in this methodology, which are easier to understand, are given bellow.

Adaptability. refers to the capacity of genotypes to utilize the environmental stimulus with advantage. As for adaptability, they are assorted into:

a) Genotypes with general or broad adaptability: in these, β_{1i} equals 1.

b) Genotypes with adaptability specific to favorable environments: in these, β_{1i} is greater than 1.

c) Genotypes with adaptability specific to unfavorable environments: in these, β_{1i} is lesser than 1.

Stability. refers to the capacity of genotypes to display a highly predictable behavior due to the stimulus from the environment. It is evaluated by the variance component attached to the regression deviations σ^2_{di}. The following types of genotypes are found:

a) Genotypes with high stability of predictability: in these, σ^2_{di} equals 0.

b) Genotypes with low stability or predictability: in these, σ^2_{di} is greater than 0.

Eberhart and Russell (1966) considered that the ideal genotype is the one which has a high production average, regression coefficient equal to 1 and regression deviations as small as possible.

Estimation of stability and adaptability parameters

Parameters β_{0i} and β_{1i} are estimated through the following expressions:

$$\hat{\beta}_{0i} = \bar{Y}_{i.} \quad \text{and} \quad \hat{V}(\hat{\beta}_{0i}) = \frac{1}{a}\hat{\sigma}^2_{\varepsilon} \quad \text{and}$$

$$\hat{\beta}_{1i} = \frac{\sum_j Y_{ij} I_j}{\sum_j I_j^2} \quad \text{and} \quad \hat{V}(\hat{\beta}_{1i}) = \frac{1\ \hat{\sigma}^2_{\varepsilon}}{\sum_j I_j^2}$$

where

$$I_j = \frac{1}{g}\sum_i Y_{ij} - \frac{1}{ag}\sum_i \sum_j Y_{ij} \quad \text{and}$$

$$\hat{\sigma}^2_{\varepsilon} = \frac{1}{r}\hat{\sigma}^2 = \frac{QMR}{r}$$

The hypothesis $H_o : \beta_{1i} = 1$ versus $H_a : \beta_{1i} \neq 1$ is evaluated by the statistic t, given by

$$t = \frac{\hat{\beta}_{1i} - 1}{\sqrt{\hat{V}(\hat{\beta}_{1i})}}$$

The stability parameter σ^2_{di} is estimated by the variance analysis method, based on the mean square of the regression deviation for each genotype (MSDi) and on the mean square of the residue, that is:

$$\hat{\sigma}^2_{di} = \sum_j \hat{\delta}^2_{ij} / (a - 2) = \frac{MSD_i - QMR}{r}$$

where

$$MSD_i = \frac{r}{a - 2}\left[\sum_j Y_{ij} - \frac{Y_i^2}{a} - \frac{\left(\sum_j Y_{ij} I_j\right)^2}{\sum_j I_j^2}\right] \qquad \text{(valid for every } i\text{)}$$

Sometimes, it may occur that many genotypes with superior average of yield are presented σ^2_{di} statistically different than zero. However, selection of a few genotypes from the group in which stability (or predictability) is low may be necessary. In these cases, an auxiliary measure of comparison between these genotypes is the determination of coefficient R_i^2, given by

$$R_i^2 = \frac{SS(Linear\ regression)_i}{SS(A/G_i)} \times 100$$

To select alfalfa genotypes by adaptability and stability of the "dry matter" trait, Vasconcelos et al. (2008) evaluated 92 genotypes in two periods, rain season and dry season. In this study, the traditional methodology cited by Cruz et al. (2004) was used, in addition to the one by Eberhart and Russell (1966) and the centroid one (Rocha et al., 2005). The result regarding data analysis through the traditional method indicated that genotype WL 612, with the smallest mean square value for environments within the genotype, was the one with less variation in the mean of cuttings in the three environments and with the greatest stability. However, this genotype had a low average of dry matter production, in comparison to the others. According to Cruz et al. (2004), it is very likely that genotypes with smaller mean square of environments within genotypes had a reduced mean. The result of the adaptability and stability analysis, by the Eberhart and Russel methodology, indicated that genotype LE

N 4 was the one presenting a high value of $\hat{\beta}_0$ (or mean equal to 1,834.83 kg of dry matter per cut per ha), and $\hat{\beta}_1$ equal to 1.015, but the estimated regression deviation for this genotype was greater than zero, indicating the material has low stability. Lédo et al. (2005) also found this behavior, reporting that, according to Cruz et al. (2004), it may sometimes occur for several cultivars with high average yield, to present regression deviations statistically different than zero. In this case, cultivars from the group that showed reduced stability can be selected, but using the value of *R2* as an auxiliary measure. Vasconcelos et al. (2008) also highlighted that the Eberhart and Russel methodology could be more efficient if the number of environments was greater. Since the number of environments was three, the number of points used for regression was also three. By the centroid method, 36 out of the 92 genotypes were deemed as having general adaptability, but among them the materials Crioula, LE N 4 and P 30 showed the greatest probabilities (64%, 81% and 75%, respectively) of belonging to class I, that is, of having general adaptability to the environments. The genotypes with best behavior in rain environments were Rocio and Costera SP INTA with 35.7% and 35.0% probabilities of belonging to that class. After evaluation by the three methods, the authors concluded that genotype LE N 4 has general adaptability (centroid method) and high mean of production, and that it is responsive to environmental improvement for dry matter production (Eberhart and Russel method). Genotypes P 5715 and Bárbara SP INTA had, in addition to good adaptability, stability in dry matter production (centroid method). Genotype Bacana had the best adaptability to dry environments (centroid method), with an acceptable mean of dry matter production and of stability (traditional method), and thus constitutes a good option for forage exploitation throughout the year.

When choosing a genotype, it is expected for its initial superiority to last throughout its life. Similarly, it is also expected that the good performance presented in certain structures or integral parts of the individual reflects the potential of the genotype to be used as a whole. The truth of this expectation can be proved by the repeatability coefficient of the characteristic studied. The repeatability coefficient can be obtained when the measurement of a given trait is performed repeatedly to the same individual in time or in space. Generally, this coefficient is useful for breeding, because it allows evaluating the number of necessary measurements to have a good prediction of the real value of the individual. Also, repeatability represents the maximum value that can be reached by heritability.

Different methodologies can be used to obtain the repeatability coefficient. The method presented bellow is based on the variance analysis, according to the model:

$$Y_{ij} = \mu + g_i + a_j + \varepsilon_{ij}$$

where

Y_{ij} : observation regarding the i-th genotype in the j-th environment (time or space)

μ : general mean

g_i : random effect of the i-th genotype under influence of the permanent environment (i = 1, 2, ..., p)

a_j : steady effect of the temporary environment on the j-th measurement (j = 1,2,..., η) and

ε_{ij} : experimental error set by the temporary environmental effects on the j-th measurement of the i-th genotype

The scheme of the variance analysis for the model with two variation factors (g and a) is presented in Table 10.

When we evaluate p genotypes in η repeated measurements, we can estimate the repeatability coefficient by the intraclass correlation obtained from the variance analysis. The repeatability coefficient is given by:

$$r = \hat{\rho} = \frac{C\hat{o}v(Y_{ij}\ ,\ Y_{ij'})}{\sqrt{\hat{V}(Y_{ij})\ \hat{V}(Y_{ij'})}} = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_Y^2} = \frac{\hat{\sigma}_g^2}{\hat{\sigma}^2 + \hat{\sigma}_g^2}$$

The determination or the precision for predicting the real value of the individual based on the mean of η evaluations is given by:

$$R^2 = \frac{\eta\rho}{1 + \rho(\eta - 1)}$$

Table 10. Scheme of the variance analysis for the model with two variation factors.

VF	GL	MS	E(MS)
Genotypes	p-1	MSG	$\sigma^2 + \eta\sigma_g^2$
Environments	a-1	MSE	-
Residue	(p-1)(a-1)	MSR	σ^2

Another interesting calculation is the prediction of the number of measurements (η_0) necessary to achieve a given level of precision (or determination), in the comparison of genotypes, for a given characteristic whose repeatability coefficient *r* is known. The expression to obtain the number of measurements is:

$$\eta_0 = \frac{R^2(1 - r)}{(1 - R^2)r}$$

Low value of repeatability coefficient indicates that there has not been regularity on the repetition of the trait from one measurement to the other; with that we should not perform reductions in the number of measurements performed to save time and labor.

In a study to obtain the repeatability estimates for dry matter production in alfalfa, Souza-Sobrinho et al. (2004) used productivity data obtained in four assays of evaluation of alfalfa carried out in different regions of the state of Minas Gerais. The authors estimated the repeatability coefficient for each of the assays, through the variance analysis method, the principal components method based on the covariance and correlation matrix and through the structural analysis based on the correlation matrix. The average repeatability estimate for dry matter production in the four assays evaluated was 0.59, with average determination coefficient of 0.96. They also concluded that carrying out an average of only four cuts was enough to learn the real genotypic value of the cultivars tested, with 85% reliability.

Ferreira et al. (1999) evaluated 42 cultivars and estimated the repeatability coefficient of the characteristics dry matter production, crude protein content in leaves and in stem and disease tolerance, assessed in rain and dry seasons, in six cuts. The authors found that the repeatability coefficient generally showed low dimension estimates (under 0.4). As for dry matter production, the repeatability coefficient ranged between 0.3195 and 0.4270, the genotypic determination was around 65% and the possibility of reaching prediction of the real value was through seven to nine cuts.

Final considerations

The increase in agricultural productivity associated to nutritional quality of alfalfa can be achieved through improvements in environmental conditions or in the genetic potential of individuals or populations. In many situations, genetic improvement is the only way to achieve this goal, in addition to the advantage of promoting hereditary changes. Due to the existence of great

genetic diversity in alfalfa, it becomes possible to select and recombine genetic forms that are more adapted, more efficient and that have better quality.

Using biometrics during the stages of a breeding program for alfalfa becomes an extremely useful tool for the researcher to make decisions, allowing the gathering of the maximum information on the evaluated experiments. With that, the strategies to conduct the next stages of the breeding program can be planned, increasing the gains achieved and the success in improving the crop.

References

BOLANÕS-AGUILAR, E. D.; HUYGHE, C.; DJUKIC, D.; JULIER, B.; ECALLE, C. Genetic control of alfalfa seed yield and its components. **Plant Breeding**, v. 120, p. 67-72, Apr. 2001.

CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. 3. ed. Viçosa: Ed. da UFV, 2004. 480 p.

CRUZ, C. D. **Princípios de genética quantitativa**. Viçosa, MG: Ed. da UFV, 2005. 391 p.

EBERHART, S. A. Factors affecting efficiencies of breeding methods. **African Soils**, v. 15, p. 669-680, 1970.

EBERHART, S. A.; RUSSEL, W. A. Stability parameters for comparing varieties. **Crop Science**, v. 1, n. 5, p. 36-40, Jan. 1966. DOI: https://doi.org/10.2135/cropsci1966.0011183X000600010011x.

EDWARDS, A. W. F.; CAVALLI-SFORZA, L. L. A method for cluster analysis. **Biometrics**, v. 21, p. 362-375, 1965.

FERREIRA, R. P.; BOTREL, M. A.; PEREIRA, A. V.; CRUZ, C. D.; Avaliação de cultivares de alfafa e estimativas de repetibilidade de caracteres forrageiros. **Pesquisa Agropecuária Brasileira**, v. 34, n.6, p. 995-1002, June 1999. DOI: https://doi.org/10.1590/S0100-204X1999000600010.

GRIFFING, B. Concept of general and specific combining ability in relation to diallel crossing systems. **Australian Journal of Biological Science**, v. 9, p. 463-493, 1956. DOI: https://doi.org/10.1071/BI9560463.

LÉDO, F. J. S.; BOTREL, M. A.; EVANGELISTA, A. R.; VIANA, M. C. M, PEREIRA, A. V.; SOBRINHO, F. S.; OLIVEIRA, J. S.; XAVIER, D. F.; HEINEMANN, A. B. Adaptabilidade e estabilidade de cultivares de alfafa avaliadas em Minas Gerais. **Ciência e Agrotecnologia**, v. 29, n. 2, p. 409-414, abr. 2005. DOI: https://doi.org/10.1590/S1413-70542005000200019.

MADRIL, C. M.; PIERCE, C. A.; RAY, I. M. Heterosis among hybrids derived from genetically improved and unimproved alfalfa germplasm. **Crop Science**, v. 48, p. 1787-1792, Sept. 2008. DOI: https://doi.org/10.2135/cropsci2008.01.0050.

RAO, C. R. **Advanced statistical methods in biometric research**. New York: Wiley, 1952. 400 p.

ROCHA, R. B.; ABAD, J. I. M.; ARAÚJO, E. F.; CRUZ, C. D. Avaliação do método centróide para estudo de adaptabilidade ao ambiente de clones de *Eucalyptus grandis*. **Ciência Florestal**, v. 15, n. 3, p. 255-266, July/Sept. 2005. DOI: https://doi.org/10.5902/198050981863.

SEGOVIA-LERMA, A.; CANTRELL, R. G.; CONWAY, J. M.; RAY, I. M. AFLP-based assessment of genetic diversity among nine alfalfa germplasms using bulk DNA templates. **Genome**, v. 46, p. 51-58, Feb. 2003. DOI: https://doi.org.10.1139/g02-100.

SOUZA-SOBRINHO, F.; LÉDO, F. J. S.; PEREIRA, A. V.; BOTREL, M. A.; EVANGELISTA, A. R.; VIANA, M, C. M; Estimativas de repetibilidade para produção de material seca em alfafa. **Ciência Rural**, v. 34, n. 2, p. 531-537, Apr. 2004. DOI: https://doi.org/10.1590/S0103-84782004000200030.

TOUIL, L.; GUESMI, F.; FARES, K.; FERCHICHI, A. Genetic diversity of some Mediterranean populations of the cultivated alfafa (*Medicago sativa L.*) using ISSR markers. **Biotechnology**, v. 7, n. 4, p. 808-812, 2008. DOI: https://doi.org/10.3923/biotech.2008.808.812.

VASCONCELOS, E. S.; JÚNIOR, W. B.; CRUZ, C. D.; FERREIRA, R. P.; RASSINI, J. B.; VILELA, D. Seleção de genótipos de alfafa pela adaptabilidade e estabilidade da produção de matéria seca. **Acta Scientiarum, Agronomy**, v. 30, n. 3, p. 339-343, set. 2008. DOI: https://doi.org/10.4025/actasciagron.v30i3.3511.

WARD, J. H. Hierarchical grouping to optimize an objective function. **Journal of the American Statistical Association**, v. 58, p. 236-244, 1963.