Maydica 54 (2009): 269-285

'FANDANGO': LONG TERM ADAPTATION OF EXOTIC GERMPLASM TO A PORTUGUESE ON-FARM-CONSERVATION AND BREEDING PROJECT

P.M.M. Mendes-Moreira[1,2,9,*], M.C. Vaz Patto[2,9], M. Mota[3], J. Mendes-Moreira[4,5], J.P.N. Santos[1], J.P.P. Santos[1], E. Andrade[6], A.R. Hallauer[7], S.E. Pego[8,9]

[1] *Escola Superior Agrária de Coimbra, Departamento de Fitotecnia, Sector de Protecção Vegetal, Portugal*
[2] *Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Portugal*
[3] *Estação Agronómica Nacional, Instituto Nacional de Recursos Biológicos, Portugal*
[4] *Faculdade de Engenharia da Universidade do Porto, DEI, Portugal*
[5] *LIAAD-INESC Porto L.A., Portugal*
[6] *Banco Português de Germoplasma Vegetal, Instituto Nacional de Recursos Biológicos, Portugal*
[7] *Faculty of Agronomy, Iowa State University, Ames, IA 50010, USA*
[8] *Fundação Bomfim. Rua da Boavista, 152-154, 4700-416 Braga, Portugal*
[9] *Zea+, Portugal*

Received July 6, 2009

ABSTRACT - Climatic change emphasize the importance of biodiversity maintenance, suggesting that germplasm adapted to organic, low input, or conventional conditions is needed to face future demands. This study presents:
I - The two steps genesis of the synthetic maize population 'Fandango'. A) 'NUTICA' creation: in 1975, Miguel Mota and Silas Pego, initiated a new type of polycross method involving 77 yellow elite inbred lines (dent and flint; 20% Portuguese and 80% North American germplasm) from the NUMI programme (NUcleo de melhoramento de MIlho, Braga, Portugal). These inbreds were intermated in natural isolation and progenies submitted to intensive selection for both parents during continued cycles; B) From 'NUTICA' to 'Fandango': 'Fandango' was composed of all the crosses that resulted from a North Carolina Design 1 matting design (1 male crossed with 5 females) applied to 'NUTICA'.
II - The diversity evolution of 'Fandango' under a Participatory Breeding project at the Portuguese Sousa Valley region (VASO) initiated in 1985 by Pego, with CIMMYT support. Morphological, fasciation expression, and yield trials were conducted in Portugal (3 locations, 3 years) and in the USA (4 locations, 1 year) using seeds obtained from five to seven cycles of mass selection (MS). The selection across cycles was done by the breeder (until cycle 5) and farmer (before cycle 11 till present). ANOVA and regression *analysis* on the rate of direct response to selection were performed when the assumption of normality was positively confirmed. Otherwise the non parametric Multivariate Adaptive Regression Splines (MARS) was performed.
Response to mass selection in Iowa showed significant decrease in yield, while in Portugal a significant increase for time of silking, plant and ear height, ear diameters 2, 3, 4, kernel number, cob diameters, and rachis was observed. At this location also a significant decrease was observed for thousand kernel weight and ear length. These results showed that mass selection were not effective for significant yield increase, except when considered Lousada with breeder selection (3.09% of gain per cycle per year). Some non-parametric methods (MARS, decision trees and random forests) were used to get insights on the causes that explain yield in Fandango. Kernel weight and ear weight were the most important traits, although row numbers, number of kernels per row, ear length, and ear diameter were also of some importance influencing 'Fandango' yield.

KEY WORDS: *Zea mays*; 'Fandango'; Open-pollinated varieties; Synthetic variety; Participatory plant breeding; Genetic diversity.

Abbreviations
NUMI - Maize Breeding Station (Núcleo de Melhoramento de Milho)
OPV - Open-pollinated Variety
PMB - participatory maize breeding
VASO - Sousa Valley, Portugal
MARS -Multivariate Adaptive Regression Splines
CART - Classification and Regression Trees
RF - Random Forests.

* For correspondence (pmoreira@esac.pt).

INTRODUCTION

Sustainability in agriculture emphasises the need for organic and low input systems. This suggests that older varieties, landraces, and synthetics, typical from these systems, could provide materials for use in marginal areas and supply breeding programs with germplasm that could be useful in different agriculture practices and systems (e.g. rotation and polycropping systems) (TILMAN *et al.*, 2002; WOLFE *et al.*, 2008).

HALLAUER (1994) proposed four distinct stages for maize breeding: 1) domestication; 2) development of maize races by Native Americans till 16^{th} century; 3) development of varieties from the original races by American and European colonists (1500 till 1925) and 4) development of inbreds and hybrids (1909 till present). Overlaps can occur between these stages. Portuguese maize history includes stage 3 and 4.

In Portugal, stage three begun after the discovery of the Americas by Columbus (1492) (FERRÃO, 1992). Maize was responsible for shaping the landscape (e.g., terraces, water mills, and store facilities), people (e.g., traditions, religion, language and standard of living), the economy (e.g., maize as payment to landlords), and type of food (e.g., directly for maize bread and indirectly through meat consumption). The impact of the maize expansion from the Southern Portuguese region of Algarve to the Northwest areas of the country led to genetic adaptation to a diversified number of microclimates, according to the sequence of valleys and mountains in these regions (PEGO and ANTUNES, 1997; MOREIRA, 2006). This stage in the Northwest still continues through *on-farm* conservation (VAZ PATTO *et al.*, 2007) and participatory maize breeding.

Stage four started in Portugal after World War II, when the USA success in maize breeding had a tremendous impact in Europe because of the availability of hybrid seed. North American hybrids were tested across Europe and trials in Portugal were successful. Breeding stations were established within Portugal, from North to South in the cities of Braga (NUMI), Porto, Viseu, Elvas and Tavira. Nevertheless, adoption of American maize hybrids did not succeed at that time, because hybrids did not satisfy the farmers needs (e.g., quality for maize bread and intensified polycropping systems). On these maize breeding stations, inbreds primarily from Portuguese and American germplasm sources were developed and based on these new inbreds, hybrids were made and tested. NUMI was responsible for the overall national program and the production of national important hybrids (e.g., HB3/BRAGA).

In 1984, Silas Pego started, with the CIMMYT support, an *on-farm* participatory maize breeding (PMB) project at the Portuguese Sousa Valley region (VASO). VASO was intended to answer the needs of small farmers (e.g., yield, bread making quality, ability for polycropping systems) with scarce land availability due to a high demographic density, where the American agriculture model did not fit and the multinationals had no adequate market to operate. To implement this project an *integrant philosophy* approach was developed (PEGO and ANTUNES, 1997; MOREIRA, 2006) and three main decisions were made: 1) the choice of the location to represent the region, 2) the farmer to work with, side-by-side (considering the farmer as the most important genetic resource where the decision power resides; i.e., respecting the "system" would imply accepting low input and intercropping characteristics, as well as accepting and respecting the local farmer as the decision maker) and 3) the germplasm source (PEGO and ANTUNES, 1997; MOREIRA, 2006). This breeding project was applied to local landraces (e.g., 'Basto', 'Aljezur', 'Aljezudo', 'Castro Verde', 'Verdial de Aperrela' and 'Verdial de Cete', 'Amiúdo' and 'Pigarro') (MOREIRA, 2006), and to a synthetic population 'Fandango'. The 'Fandango' represents a transversal project between *on-station* and *on-farm* programs, which means also the overlapping between third and fourth stage; i.e., adaptation to farmers needs through participatory maize breeding and on-station breeding programs.

Objectives of our study were to summarize research on: 1) the adaptation and evolution of the exogenous synthetic population 'Fandango' during 22 years of mass selection by breeder and farmer; 2) to determine the more representative traits related with yield, that could be useful for future selection; and 3) The "Sousa Valley Best Ear" competition and its relationship with 'Fandango' and participatory plant breeding.

MATERIALS AND METHODS

The germplasm

'NUTICA' - The 'NUTICA' (FAO 700) is the acronym of NUMI (maize breeding centre in Portugal) and Departamento de GenéTICA-EAN (Department of Genetics) and represents a maize synthetic according to the definition of LONNQUIST (1961). In 1975, after one year of material preparation, Miguel Mota (MOTA *et al.*, 1978) and Silas Pego, initiated a new type of polycross method involving 77 yellow, elite inbred lines (dent and flint; 20% Portuguese and 80% American germplasm) from the NUMI programme.

The 77 inbreds were intermated in natural isolation (from other maize) and progenies submitted to intensive selection among parents during continued cycles from 1975 to 1978. The

TABLE 1 - *Mass selection applied to Fandango since 1985, selected cycles for trials evaluation (locations and years), seasons per cycle and standard populations used.*

Selection method																							**Year.Cycle-1**
Year	1985	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	04	05	06	07	
Cycles of:																								
Mass Selection	C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	C21	C22	1

Evaluation trials									**Standard populations**	
Mass Selection (Breeder)		C1-86	C3-88	C5-90					BS22 and BS21(R)C9	
Mass selection (Farmer)					C11-96	C15-00	C19-04	C22-07	**TEPR-EC6*** **	
Locations (with 3 replications)										NUTICA
Iowa (2005)		4	4	4	4	4	**		4	
Portugal (2005)		3	3	3	3	3	3		3	3
Portugal (2007)		3		3	3	3	3			
Portugal (2008)		3	3	3	3	3	3	3	**3*****	3
Multiplication seed stock	2005		05	05	05	05	05		05	05

* - drought after sowing at Montemor-o-Velho location lead to data exclusion; ** - C19-04, due to seed injuries data were excluded; *** - TEPR-EC6 was included in 2008 trials; Cx-y, where C-cycle, x-number of cycles, y – year correspondent to cycle of selection; in shadow - corresponds to the time frame of selection by breeder and farmer, some of this cycles were kept in cold storage.

synthetic 'NUTICA' was then used to obtain S2 lines (1983 at ENMP) and subpopulations were constituted based on ear shape (at NUMI): 1 - 'Estica' – selection for ears with length of equal/more than 26 cm; 2 – 'Bucha' – selection for ears with equal/more than 20 kernel rows; and 3 – 'Fisga' – selection for plants with prolificacy. 'Fandango' was another sub-population also originated from 'NUTICA' as a result from the application of North Carolina matting Design 1.

'FANDANGO' - In 1983, the latest version of 'NUTICA' (almost entirely yellow dent) was included in Pego's breeding program at ENMP (Elvas Breeding Station). In 1984, with the purpose of evaluating the gene action composition (additive vs non-additive), the population was submitted to North Carolina matting Design 1 (1 male crossed with 5 females), as part of the MSc project of Fátima Quedas under Pego's supervision. The results obtained in the 2nd year trial (complete randomized design) were very promising, with higher yielding levels obtained in the borders (composed by a mixture of all crosses in the trials). Due to the isolation conditions of the field, Pego used a mixture obtained in open pollination as a first basis of what would be designated as 'Fandango'. This first bulk of seed (700 kg) was distributed to all the Portuguese departments of agriculture, from Vale do Tejo to Minho Region and micro-trials were established. The feedback received from those departments was very positive even in altitude areas either for ear size, or for yield (Pego, personal communication).

Based on the good results, in 1985 Pego introduced 'Fandango' at Lousada (Northwest of Portugal) and phenotypic recurrent selection have been applied by breeder (stratified mass selection, till cycle 5) and farmer since then (Pego and Antunes, 1997).

The introduction in 1985 was done in an area of 1 ha located in a strategic place for farmers' observation. This location permitted that two main goals were fulfilled: 1) engage farmers with the VASO project through the big ears and good yields obtained with 'Fandango' (the seed obtained was then given to farmers); and 2) provide the link between *on-station* and *on-farm* breeding purposes.

For the VASO project, 'Fandango' selection was not the main goal of the project, so less attention was given compared with 'Pigarro' (Mendes Moreira *et al.*, 2008).

The 'Fandango' is a FAO 600 population, with yellow dent kernels, is characterized for having both high kernel row numbers (between 18 and 26) and large ear size. These characteristics explain why in each of the past 17 years, Fandango has been the winner of the contest "Best ear of Sousa Valley Region" within the 'yellow dent' category.

Phenotypic recurrent selection (mass selection)

The phenotypic recurrent selection or mass selection began in 1985 at Lousada and can be divided in two phases: 1) from 1985 till 1996, selection was mainly done by the breeder; and 2) after 1996, farmer selection phase, in which the farmer was more engaged with the project.

The breeder program included two parental controls (stratified mass selection with parental control c = 1.0) and selection was conducted under a three step sequence (A - B - C):

A) immediately before the pollen shedding, selection is performed for the male parent by detasseling all the undesirable plants (pest and disease susceptible, weakest and those that do not fit the desirable ideotype);
B) before harvest, besides selecting for the best ear size, the plants are foot kicked at their base (first visible internodes) to evaluate their root and stalk quality. With this procedure, as an indirect measurement, the pest and disease tolerance can be evaluated. In practical terms, if the plant breaks, it is eliminated. A special selection preference is given to prolific plants;
C) at the storage facilities, after harvest, selection is performed separately for both normal and prolific ears and always includes ear length, kernel-row number, prolificacy, and the

elimination of damaged/diseased ears. The selected ears are shelled and mixed together to form the next generation seed. The breeder selection pressure ranged from 1 to 5%.

The farmer pursued the mass selection procedure more commonly used (for one parental control c = 0.5) and only at step C. Success has not been easy to achieve in convincing the farmer to adopt the two parental control at step A, and only partially at step (B) (Fig. 1, Table 1). The farmer selection pressure ranged from 1 to 5%.

Germplasm evaluation

Germplasm management - Since the beginning of the VASO Project, phenotypic data were collected and some seed of selection cycles of 'Fandango' was kept at 4°C at BPGV (Portuguese Plant Germplasm Bank, Braga, Portugal) cold storage facilities.

Seed of cycles C1-86, C3-88, C5-90 (obtained by the breeder) and cycles C11-96, C15-00, C19-04 and C22-07 (obtained by the farmer) of phenotypic recurrent selection, from NUMI (Table 1) were chosen and used for the trials conducted in 2005, 2007, and 2008 (in 2007, C3-88 was not included due to area limitations and in 2008 the C22-07 was included to test the new cycle of selection). In parallel, the selection cycles seed stock used in the trials, were multiplied by hand pollination in 2005, except for C22-07 seed. All pollinated ears were harvested and dried at approximately 35°C to obtain a uniform moisture level of 13 to 14%.

Evaluation trials - To determine the effectiveness of mass selection in 'Fandango', trials were conducted at several locations in Portugal and Iowa (Table 1):

1. Five to seven cycles of mass selection (breeder 2-3 cycles, farmer 3-4 cycles);
3. three replication trials for each entry and location;
4. trials conducted in four locations within Iowa-USA (Calumet, Kanawha, Ames and Nashua) during 2005 and three locations within Portugal during 2005, 2007 and 2008 (Lousada, Montemor-o-Velho, and Coimbra).

At Iowa, two row plots (5.47 m long with 0.76 m between rows) were overplanted by using a machine planter. Each plot was thinned at the seven-leaf stage to 50 plants per plot for a plant density of 60 000 plants ha^{-1}. All the plots were harvest by machine, with grain yield and grain moisture data recorded electronically on the harvester.

In Portugal, two rows plots (at Lousada 6.9 m long with 0.70 m between rows, and in the other locations 6.4 m long with 0.75 m between rows) were overplanted by hand. Each plot was thinned at the seven-leaf stage from 48 (Coimbra and Montemor-o-Velho) to 50 (Lousada) plants per plot for a stand of 50 000 plants ha^{-1}. All the plots in Portugal were harvested by hand. Plots were either mechanically and/or hand weeded as necessary.

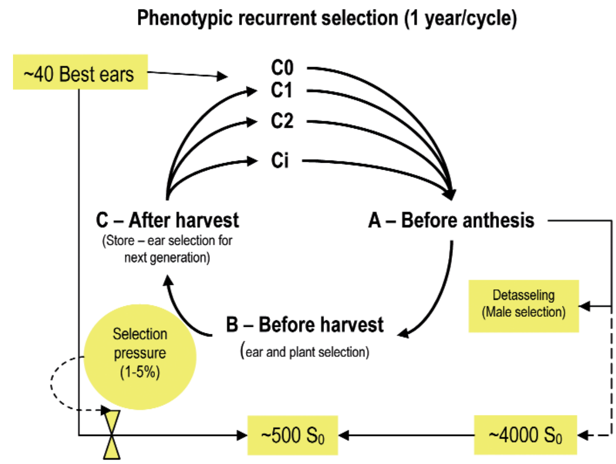

FIGURE 1 - Phenotypic recurrent selection methodology used in 'Fandango' by the breeder.

Germplasm for comparisons - The North American populations BS21(R)C9 and BS22(R)C9 (Hallauer *et al*., 2000), were included on 2005 trials, and TEPR-EC6 (Troyer, 2000) was included also on 2008 trials (Table 2). These populations were used as standards regarding the cycle of 'Fandango'. They were included to better understand the differences between USA and Portugal environments, and because these populations are better known than 'Fandango' by the international scientific community. 'NUTICA' was also included.

Data collection - Data were obtained in all the field trials for final stand, silk emergence (only Ames at Iowa), root lodging, stalk lodging and grain yield (Mg ha^{-1}) adjusted to 15% grain moisture at harvest (moisture during harvest in Portugal, was measured with a moisture meter, using a mixture sample of five shelled ears grain). These ears were also weighted, as well as the

TABLE 2 - *Traits measured per location and per plot, codes and respective description.*

Traits	Measurements Iowa	Measurements Pt	Data/plot Plot	Data/plot Pl or Ears	Codes	Scale
Grain yield (15% moisture), Mg ha^{-1}	x	x	1		Yield	**a1)** hand harvest (Portugal), Grain yield = Ear weight x (Grain weight/Ear weight) five shelled ears are used for determination of this ratio and for moisture content; **a2)** combine used (Iowa), grain yield and moisture content are directly measured; **b)** Grain yield 15% moisture = Grain yield x (100% - % moisture at harvest)/(100%-15% moisture)
Grain moisture %	x	x	1		Grain moisture	**a1)** hand harvest (Portugal), grain from five shelled ears are used for moisture determination); **a2)** combine (Iowa), moisture content are directly measured
Days-to-silk, nº †	Ames	x	1		Fi	The beginning of days to silk (from planting until 50% of the plants in the plot begin silk emergence
Days-to-silk, nº † end		x	1		Ff	The end of days-to silk (from planting until 50% of the plants in the plot finish silk emergence

TABLE 2 - *Continued.*

Traits	Measurements	Data/plot			Codes	Scale	
	Iowa	Pt	Plot	Pl or Ears			
Days-to-anthesis, nº †		x	1		Mi		The beginning of days-to anthesis, i.e., from planting until 50% of the plants in the plot start anthesis
Days-to-anthesis, nº † end		x	1		Mf		The end of days-to anthesis (from planting until 50% of the plants in the plot finish anthesis
Stand	x	x	1		Plants ha^{-1}		Thousands of plants per hectare
Overlapping index		x	1		OI		This method enables the knowledge of a population concerning the relative amount of theoretical allogamy versus autogamy
Uniformity		x	1		U	1 to 9	1 - minimum uniformity and 9 - maximum; 1-5 to populations and 6-9 to inbreds.
Leaf angle		x	1		N	1 to 9	Angle of the adaxial side of the leaf above the ear with the stalk (5=45°, <5 =<45° and >5 = >45°C)
Tassel branching		x	1		T	1 to 9	1 - absent tassel (inbreeds and hybrids) 9 - a much branched tassel (frequent in populations with abnormal fasciated ears)
Ear placement		x	1		E	1 to 9	5 - indicates that the ear is located in the middle of the plant, if <5 bellow and if >5 above the middle of the plant
Root lodging %	x	x	1		R	%	Percentage of plants leaning more than 30° from vertical
Stalk lodging %	x	x	1		S	%	Percentage of plants broken at or below the primary ear node, related with the quality of the stalk and the stalk damage caused by some insect attack
Puccinia spp.		x	1		*Puccinia* spp.	1 to 9	Evaluation on the leaves surface: 1 - symptoms absence and 9 - maximum intensity of attack
Ustilago maydis		x	1		*U. maydis*	1 to 9	Evaluation on tassel, stems and ears: 1 - symptoms absence and 9 - maximum intensity of attack
Plant height, cm		x		20	H		Plant height, from the stalk basis to the last leaf insertion before the tassel
Ear height, cm		x		20	Ear height		Ear height, from the stalk basis to the highest ear bearing node
Ear length, cm		x		20	L		Ear length
Ear diameter 1 and 3, cm		x		20	DE1, DE3		Large diameter in the 1/3 bottom and top of the ear, respectively
Ear diameter 2 and 4, cm		x		20	DE2, DE4		Small diameter in the 1/3 bottom and top of the ear, respectively (90° rotation from large diameter)
Kernel-row number 1and 2, nº		x		20	R1, R2		Row number in the 1/3 bottom and top of the ear, respectively
Fasciation		x		20	Fa	1 to 9	1 - without fasciation and 9 - maximum of fasciation
Determinated/Indeterminated		x		20	D/I		Top of the ear full of grain, case of determinated ears (2) or not, case of indeterminated ears (1), average value is calculated
Convulsion		x		20	CV	0 to 5	kernel row arrangement in the ear (0 - without convulsion, regular kernel row arrangement, 5 - maximum of convulsion, without kernel row arrangement)
Flint/Dent		x		20	F/D	1 to 9	1 - Popcorn, 2 - flint, 3 - medium flint, 4 - low flint, 5 - 50% flint and 50% dent, 6 - low dent, 7 - medium dent, 8 - high dent, 9 - sweet maize
Ear weight, g		x		20	EW		Ear weight, adjusted to 15% of grain moisture
Kernel weight, g		x		20	KW		kernel weight per ear, adjusted to 15% moisture
Cob weight/Ear weight		x		20	CW/EW		Indicates the percentage of cob weight in the ear weight
Ear % moisture		x		20	Ear moisture		Determination of % moisture content per individual ear, after drying (35°C)
Kernel dept, cm		x		20	KD		Measure of one kernel in the middle of the ear
Kernel number, nº		x		20	KNº		Kernel number per ear
Thousand kernel weight, g		x		20	SW		Thousand kernels weight at 15% moisture content
Kernel per row, nº		x		20	NC		Kernel number per row
Cob diameter 1, 3, 2 and 4 cm		x		20	DC1, 3, 2 and 4		Cob diameters **1**, **3**, **2** and **4**; similar measurements as described for DE's
Medulla 1 and 2, cm		x		20	M1, M2		Large and small length of medulla, respectively
Rachis 1 and 2, cm		x		20	Ra1, Ra2		Large and small length of rachis, respectively
Cob colour		x		20	CC		Cob colour: 1 is red and 2 is white

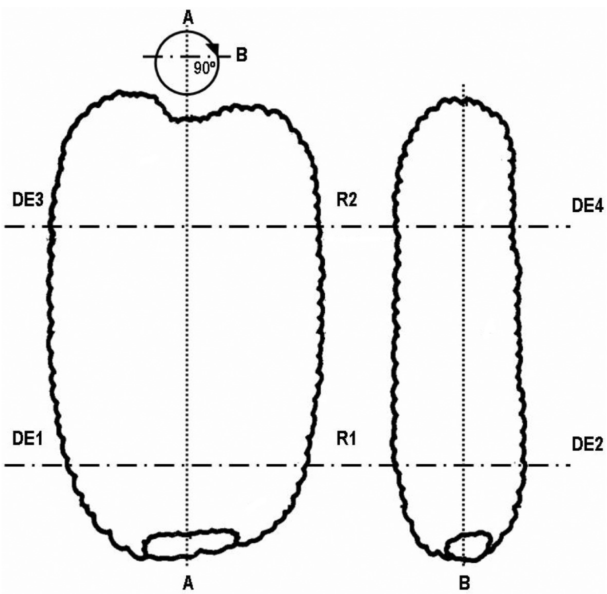

FIGURE 2 - Two orthogonal views of the same ear showing the way that the two sets of diameters and the two row numbers (R1 and R2) were measured and counted; in position A, the diameters D1 and D3 were measured; in position B (a 90° turn along the length axis), D2 and D4 were measured (adapted from PEGO and HALLAUER, 1984).

cobs, to determine the grain weight and the ratio cob/ear weight (Table 2).

For Portugal, measurements were done on plot basis or using 20 random plants or ears per plot. After harvest, the 20 random ears of each plot were dried at 35°C to approximately 15% grain moisture. Ear data included overlapping index, ear length, ear diameters, kernel-row number, ear fasciation, and other traits included in Table 2 and Figs. 2-3 (PEGO and HALLAUER, 1984; MOREIRA and PEGO, 2003; MOREIRA *et al.*, 2008).

The overlapping index determination allows prediction of the relative amount of theoretical allogamy versus autogamy of a population. The theoretical reasoning assumes that all the polinization occurs only under gravity influence, so that when a maize plant has flowering overlapping, the potential selfing will have a direct effect on the inbreeding depression. Four sets of data were collected per plot (number of days from planting to the beginning (a) or end (A) of male flowering; or to the beginning (b) or end (B) of female flowering). This data were used in the mathematical expression as follows:

$$\mathbf{OI} = \frac{\mathbf{(B - b) + (A - a) - |B - A| - |b - a|}}{\mathbf{2\,(B - b)}}$$

This formula provides information, under its own limitations, such as:
overlapping index is limited to 1 (100%);
overlapping index is either positive (some overlapping) or negative (overlapping does not occur).

Data analysis - ANOVA, linear regression and MARS. A regression analysis was conducted separately for Portuguese locations (22 cycles) where Lousada was also considered per se and Iowa locations (15 cycles) when the assumption of normality was positively confirmed. Since linear regression assumes normality, the Kolmogorov-Smirnov (KS) - variant Lilliefors (LILLIEFORS, 1967) hypothesis test was performed for each dependent variable using a Type I error of 5%. The p-value for each one of the tests is computed using the function Lillie.test from the R-project (R DEVELOPMENT CORE TEAM, 2008).

Those variables that, according to the KS-Lilliefors test, did not have a normal distribution were analyzed using a non-parametric method: MARS - Multivariate Adaptive Regression Splines (FRIEDMAN, 1991). This method was chosen because it has no assumptions and has good interpretability (HASTIE *et al.*, 2001). MARS is quite similar to stepwise regression but the relations between each dependent variable and the independent one do not need to be linear, because each one of those relations is defined by a set of connected linear segments, instead of a single one. Like linear regression, MARS result is expressed as an equation typically a bit more complex than linear regression but equally interpretable. MARS was used as many times as the number of non-normal independent variables. At each time just one variable is used. In all the experiments the dependent variable is the selection cycle. The results were obtained using the function earth from the R-project (R DEVELOPMENT CORE TEAM, 2008).

All experiments were analyzed as randomized complete block designs, with three replications. When normality (KS-Lilliefors) and homogeneity (Levene Test) were positively confirmed, analysis of variance were calculated for selection cycles,

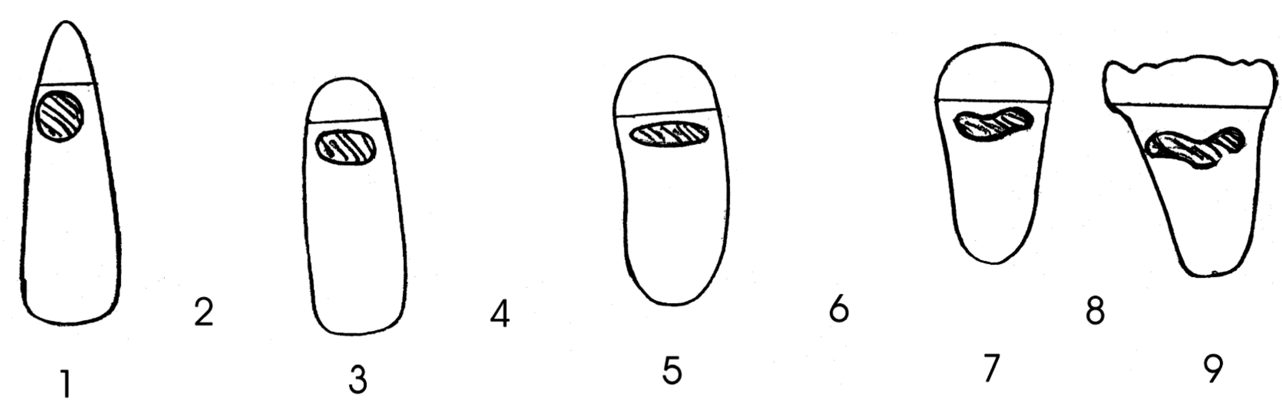

FIGURE 3 - Fasciation degree (1 - without fasciation and 9 - maximum of fasciation), shape of the ear and from transversal cut view.

TABLE 3 - *MARS for the rejected null hypothesis of normality when KS Lilliefors was used. Mean traits for standard populations.*

Iowa (C1-C15)	MARS - MULTIVARIATE ADAPTIVE REGRESSION SPLINES		Populations Standard (Iowa)			
Traits	R^2	equation: Explaining each variable along cycles	NuticaC077	BS21(R)C9	BS22(R)C9	TEPREC6
Days-to-silk, nº † end (IAmes)		**73.33333333**	82.00	73.67	73.67	73.67
Root lodging %	0.00	**0.3983491**	0.34	0.23	0.31	0.26
Stalk lodging %	0.00	**0.08103**	0.12	0.07	0.10	0.07

Portugal (C1-C22)	MARS - MULTIVARIATE ADAPTIVE REGRESSION SPLINES		Populations Standard (Portugal)			
Traits	R^2	equation: Explaining each variable along cycles	Nutica C077	BS21(R)C9	BS22(R)C9	TEPREC6
Grain moisture %	0.09	**26.4380218**+0.2993786*max{0,(Cycle-5)}	27.65	28.01	26.41	26.79
Days-to-silk, nº †	0.33	**74.1428571**+0.7173712*max{0,(Cycle-5)}-0.5814179*max{0,(Cycle-11)}	72.93	63.56	63.56	65.07
Days-to-anthesis, nº †	0.25	**72.3968254**+0.5265331*max{0,(Cycle-5)}-0.4089340*max{0,(Cycle-11)}	71.00	62.11	62.33	64.07
Days-to-anthesis, nº † end	0.25	**78.5730654**+0.3360070*max{0,(Cycle-5)}	77.20	65.33	65.67	68.53
Overlap **i**ndex	0.07	**0.67362613**-0.01893501*max{0,(Cycle-5)}+0.04105288*max{0,(Cycle-15)}	0.74	0.32	0.42	0.50
Uniformity	0.03	**2.7086093**+0.1712043*max{0,(Cycle-19)}	3.11	3.89	3.89	3.61
ANgle	0.00	**5.11875**	5.00	5.11	4.89	4.33
Tassel	0.10	**5.98704302**+0.06100515*max{0,(Cycle-11)}	5.44	4.11	4.11	4.87
Ear placement	0.00	**5.0375**	4.56	4.33	4.22	4.33
Root lodging %	0.00	**0.0420625**	0.00	0.00	0.00	0.01
Stalk lodging %	0.00	**0.0620625**	0.02	0.00	0.00	0.01
Ustilago maydis	0.00	**1.0125**	1.33	1.00	1.00	1.00
Puccinia spp.	0.00	**3.20625**	2.67	3.00	3.11	2.33
Ear length, cm	0.21	**21.5606061**-0.3997987*max{0,(Cycle-11)}+0.3713790*max{0,(Cycle-15)}	20.89	14.62	16.81	14.79
Ear diameter 1, cm	0.71	**5.17615731**+0.04274961*max{0,(Cycle-5)}+0.03960626*max{0,(Cycle-15)}-0.12634448*max{0,(Cycle-19)}	5.35	4.73	4.73	4.43
Kernel-row number 1, nº	0.86	**15.2736111**+0.2311062*max{0,(Cycle-5)}+0.2196642*max{0,(Cycle-11)}-0.9469818*max{0,(Cycle-19)}	16.74	16.24	14.80	14.80
Kernel-row number 2, nº	0.85	**15.1184534**+0.2687045*max{0,(Cycle-5)}+0.2895361*max{0,(Cycle-15)}-1.0784686*max{0,(Cycle-19)}	16.43	15.18	14.57	14.44
Fasciation	0.49	**1.5068138**+0.1425356*max{0,(Cycle-11)}-0,2156426*max{0,(Cycle-19)}	1.77	1.11	1.07	1.07
Determinated/Indeterminated	0.04	**1.094236958**-0.003544104*max{0,(Cycle-5)}	1.10	1.06	1.02	1.16
Convulsion	0.38	**1.48181818**+0.08639462*max{0,(Cycle-11)}-0.06227005*max{0,(Cycle-15)}	1.55	1.71	1.46	1.41
Cob/Ear weigth	0.05	**0.1530500504**-0.0006764187*max{0,(Cycle-11)}	0.15	0.13	0.14	0.14
Ear Moisture %	0.09	**18.2642032**+0.81369153*max{0,(Cycle-19)}-0.08858102*max{0,(19-Cycle)}	17.07	15.69	15.75	15.55
Kernel dept, cm	0.11	**1.22122357**+0.00983937*max{0,(Cycle-15)}	1.25	1.19	1.10	1.15
Kernel per row, nº	0.15	**39.4155950**-0.2027736*max{0,(Cycle-11)}+0.1601819*max{0,(11-Cycle)}	40.68	28.67	32.38	30.88
Cob diameter 3, cm	0.70	**2.91232241**+0.05636318*max{0,(Cycle-11)}-0.01976735*max{0,(11-Cycle)}-0.08929097*max{0,(Cycle-19)}	2.81	2.34	2.53	2.17
Medulla 1, cm	0.57	**1.27251066**+0.03295878*max{0,(Cycle-5)}	1.35	0.94	1.15	0.96
Medulla 2, cm	0.35	**1.20794969**+0.07290566*max{0,(Cycle-19)}-0,01030598*max{0,(19-Cycle)}	1.10	0.78	0.98	0.81
Cob colour	0.00	**1.396812**	1.47	1.91	2.00	2.00

The MARS equation contains the value of the original cycle (mean trait in bold) plus the transformation.

environments (locations), years (Iowa 05; Portugal all locations and Lousada per se 05, 07-08) and respective combinations. The same *analysis* were performed for 2 subgroups based on Iowa and Portuguese locations (all locations and Lousada per se). When significant differences were detected, post-hoc comparisons with Sheffe test were performed.

Response to mass selection for several traits was evaluated for Iowa, Portugal and Lousada using the linear regression model by regressing observed populations means on cycle of selection (b = regression of trait on cycle of selection and response was expressed relative to the C0 population, and on a year bases) or MARS.

TABLE 4 - *After positive assumption of normality, linear regression was used. Estimation of linear regression coefficient (b), their standard errors, initial cycle prediction ($\hat{C}0$), coefficients of determination (R^2) and % of gain per year (%Gain/Y) for mass selection (22 cycles in Portugal and 15 cycles in Iowa). For Iowa 5 traits were analysed and for Portugal 46 during 2005, 2007 and 2008. Mean traits for standard populations are also included.*

	Mass selection Iowa										Populations Standard Iowa			
Traits	b	$\hat{C}0$	R^2	%C/Y	C	E	Y	CxE	CxY	CxExY	NuticaC077	BS21(R)C9	BS22(R)C9	TEPREC6
Yield, Mg ha-1	-0.15 ± 0.04 *	5.33	0.84	-2.87	**			**			5.46	6.51	6.66	6.17
Grain moisture %	0.02 ± 0.02	21.62	0.33	0.11		**					20.63	18.04	20.58	17.43
Stand (Plants ha-1) ‡		54827									63525	62923	62322	62723

	Mass selection Portugal										Populations Standard Portugal			
Traits	b	$\hat{C}0$	R^2	%C/Y	C	E	Y	CxE	CxY	CxExY	NuticaC077	BS21(R)C9	BS22(R)C9	TEPREC6
Yield, Mg ha-1	-0.03 ± 0.01	8.66	0.56	-3.93						**	9.20	6.84	6.85	7.43
Days-to-silk, nº † end	0.32 ± 0.08 **	79.25	0.78	0.41	**	**	**			**	78.87	70.44	68.44	69.87
Plant height, cm	1.45 ± 0.24 **	258.40	0.88	0.56		**	**			**	261.76	216.18	210.46	199.86
Ear height, cm	1.54 ± 0.20 **	138.06	0.92	1.12	**	**	**			**	144.18	109.15	96.67	99.54
Ear diameter 3, cm	0.04 ± 0.01 **	4.50	0.88	0.85	**	**	**	**	**	**	4.75	4.23	4.21	3.96
Ear diameter 2, cm	0.02 ± 0.00 **	4.93	0.95	0.51	**	**	**	**	**	**	5.14	4.62	4.63	4.34
Ear diameter 4, cm	0.02 ± 0.00 **	4.47	0.89	0.55	**	**	**	**	**	**	4.63	4.11	4.12	3.90
Flint/Dent	0.00 ± 0.01	6.42	0.01	0.04	**	**	**	**		**	6.48	7.02	6.74	6.67
Ear weight, g	0.22 ± 0.29	269.61	0.10	0.08	**	**	**	**	**	**	274.43	156.52	172.89	147.72
Kernel weight, g	0.25 ± 0.25	227.90	0.17	0.11	**	**	**	**	**	**	234.08	135.56	148.32	126.85
Kernel number, nº	4.22 ± 1.22 *	576.14	0.71	0.73	**	**	**	**	**	**	637.04	417.57	453.82	427.72
Thousand kernel weight, g	-2.01 ± 0.55 *	397.29	0.73	-0.51	**	**	**	**	**	**	370.79	327.01	326.02	296.48
Cob diameter 1, cm	0.03 ± 0.00 **	3.26	0.96	0.95	**	**			**	**	3.44	2.93	3.12	2.66
Cob diameter 2, cm	0.02 ± 0.00 **	3.07	0.98	0.52	**	**	**	**	**	**	3.16	2.75	2.96	2.53
Cob diameter 4, cm	0.01 ± 0.00 **	2.62	0.88	0.52	**	**	**	**	**	**	2.67	2.21	2.45	2.11
Raquis 1, cm	0.03 ± 0.00 **	2.35	0.97	1.15	**	**	**	*	**	**	2.47	2.02	2.16	1.93
Raquis 2, cm	0.01 ± 0.00 **	2.11	0.77	0.54	**	**	**	**	**	**	2.16	1.80	1.93	1.73
Stand (Plants ha-1) ‡		47821									51185	50955	51875	51407

* - Significant at 0.05 probability levels; ** - Highly significant at 0.01 probability levels; † Number of days from date of planting to date of flowering; ‡ - the stand correspond to the average of the correspondent cycles.
%Gain/Y – percentage of gain per year, ANOVA for C-cycles of selection, E-environment; Y-years; x-interactions; $\hat{C}$ - predicted cycle of selection, except for stand that was calculated the average. Flowering data was not measured in Lousada, Portugal in 2008. Shaded portions distinguished where *Analysis* of Variance was not done from the white portions where non significant differences were registered.

Note that for Iowa locations or Iowa plus Portugal locations, only 15 cycles of mass selection were analyzed due to C19-04 exclusion. The C19-04 was excluded because of poor germination. Number of days-to-silk was considered only at Ames (Table 2).

Yield explanation based on the other traits - A second analysis was performed to get insights on the traits more related with the yield. Three methods for analysis have been used: MARS, Classification and Regression Trees (CART) and Random Forests (RF). The reason to use three methods instead of just one is to take advantage of their complementary characteristics to better understand what influences the yield in 'Fandango'.

The CART (Breiman *et al.*, 1984) splits, at each iteration, the examples in two subsets. The split is done by choosing the variable and a value that minimizes the sum of the mean squared error of the two resulting subsets. The result of this procedure is a tree like structure where each split is defined by a rule. The interpretation of each leaf-node is obtained by the set of rules in the nodes that define that leaf-node.

RF (Breiman, 2001) is a CART based approach, belonging to the family of ensemble methods, i.e., the use of a set of methods, instead of just one, in order to accomplish its task. RF generates several CART. Each generated CART is different because the tree is trained in a subset of the original set obtained using bagging (Breiman, 1996) and using a random subset of the original subset of features at each node. The interpretation of RF can be assessed using two different metrics (adapted for regression from Kuhn *et al.*, 2008):

- Mean Decrease Accuracy (% IncMSE): It is constructed by permuting the values of each variable of the test set (the test

TABLE 5 - *MARS for the rejected null hypothesis of normality when KS-Lilliefors was used for Lousada. Mean traits for standard populations at Lousada.*

Lousada (C1-C22)	MARS - MULTIVARIATE ADAPTIVE REGRESSION SPLINES		Populations Standard (Portugal)			
Traits	R^2	equation: Explaining each variable along cycles	Nutica C077	BS21(R)C9	BS22(R)C9	TEPREC6
Days-to-silk, nº †	0.19	**78.9007634**-0.6727099*max{0,(11-Cycle)}	68.00	60.67	60.67	60.67
Days-to-silk, nº † end	0.12	**84.9592875**-0.6221374*max{0,(11-Cycle)}	73.00	66.33	65.00	64.67
Days-to-anthesis, nº †	0.13	**75.2315522**-0.4720102*max{0,(11-Cycle)}	66.67	60.00	59.67	61.33
Days-to-anthesis, nº † end	0.14	**80.8396947**-0.5225827*max{0,(11-Cycle)}	72.00	63.00	62.67	66.67
Uniformity	0.00	**2.6852**	3.00	3.67	4.00	3.67
a**N**gle	0.11	**4.87533093**-0.09241877*max{0,(Cycle-15)}	5.17	5.00	4.67	4.17
Tassel	0.17	**6.10919406**+0.07165617*max{0,(Cycle-11)}	6.00	4.00	4.00	4.83
Ear placement	0.00	**5.0556**	4.50	4.00	3.67	4.17
Root lodging %	0.00	**0.0355**	0.02	0.01	0.00	0.04
Stalk lodging %	0.00	**0.1071**	0.09	0.00	0.00	0.08
Ustilago maydis	0.00	**0.6852**	0.50	1.00	1.00	0.50
Puccinia spp.	0.00	**2.1667**	1.50	1.00	1.33	1.50
Ear diameter 1, cm	0.71	**5.41178227**+0.05377996*max{0,(Cycle-11)}-0.02942301*max{0,(11-Cycle) }	5.40	4.84	4.91	4.70
Kernel-row number 1, nº	0.79	**15.4113404**+0.3027652*max{0,(Cycle-5)}	16.98	16.40	14.93	15.63
Kernel-row number 2, nº	0.75	**15.3078161**+0.2852895*max{0,(Cycle-5)}	16.59	15.33	14.83	15.20
Determinated/indeterminated	0.00	**1.0750**	1.15	1.07	1.05	1.27
Convulsion	0.34	**1.53684231**+0.05397529*max{0,(Cycle-11)}	1.61	1.62	1.28	1.33
Kernel colour	0.00	**4.1602**	4.38	4.00	3.87	3.93
Ear moisture %	0.12	**19.7479745**-0.2049248*max{0,(15-Cycle)}	16.28	14.88	15.39	15.12
Cob colour	0.00	**1.3989**	1.45	1.90	2.00	2.00

The MARS equation contains the value of the initial cycle (mean trait in bold) plus the transformation.

set is the out-of-bag subset that results from the bagging process), recording the prediction and comparing it with the unpermutated test set prediction of the variable (normalized by the standard error). It is the average increase in squared residuals of the test set when the variable is permuted. A higher % IncMSE value represents a higher variable importance.

- Mean Decrease MSE (IncNodePurity): Measures the quality (NodePurity) of a split for every variable (node) of a tree. Every time a split of a node is made on a variable, the sum of the mean squared error (MSE) for the two descendent subsets is less than the MSE for the parent subset. Adding up the MSE decreases for each individual variable over all the generated trees gives a fast variable importance that is often very consistent with the permutation importance measure. A higher IncNodePurity value represents a higher variable importance; i.e. nodes are much ‘purer’.

RESULTS

Response to mass selection

Number of days-to-silk showed significant differences (P<0.01 and P<0.05) among selection cycles. Significant differences were also found between environments (all locations at Portugal and Iowa) for all traits in the *analysis*. The cycle x environment interaction (selection cycle x location) was significant for moisture and stand, but not for yield. Significant differences found for G x E interaction, plus the different sets of data for Iowa and Portugal and different trial conditions (e.g., stand) led us to consider Iowa and Portugal as separated groups (*analysis* not shown).

Lousada (Portuguese location) was analyzed *per se* because it represents the location where the long term *on-farm* selection occurred and because significant differences found for genotype, year and location interaction exist for the majority of traits (Table 4).

Mass Selection at Iowa - The regression analysis conducted to estimate direct response to selection revealed significant decrease for yield (Tables 3, 4). Greater proportion of the variation was explained

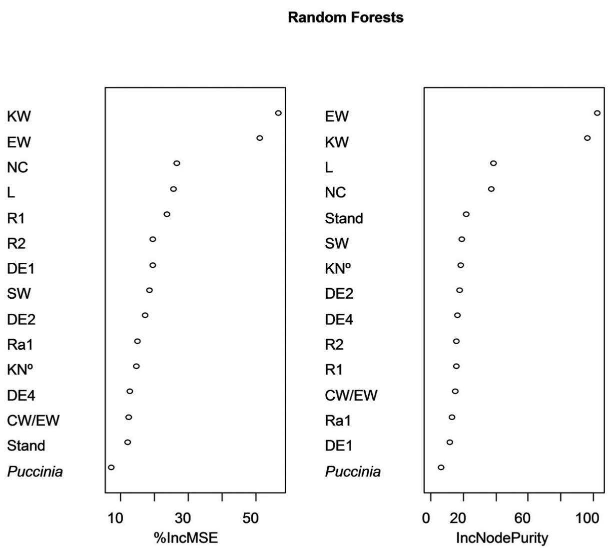

FIGURE 4 - Mean Decrease Accuracy (% IncMSE) and Mean Decrease MSE (IncNodePurity): there is no clear guidance on which measure to prefer (Kuhn *et al.*, 2008). The independent variable is Yield. They are presented only the 15 most relevant dependent variables. The percentage of variation explained was 46.7%.

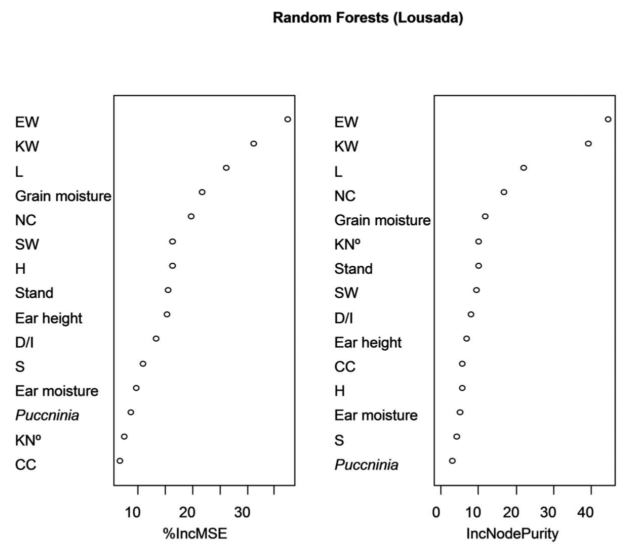

FIGURE 5 - Mean Decrease Accuracy (% IncMSE) and Mean Decrease MSE (IncNodePurity): there is no clear guidance on which measure to prefer (Kuhn *et al.*, 2008). The independent variable is Yield for Lousada. They are presented only the 15 most relevant dependent variables. The percentage of variation explained was 54.4%.

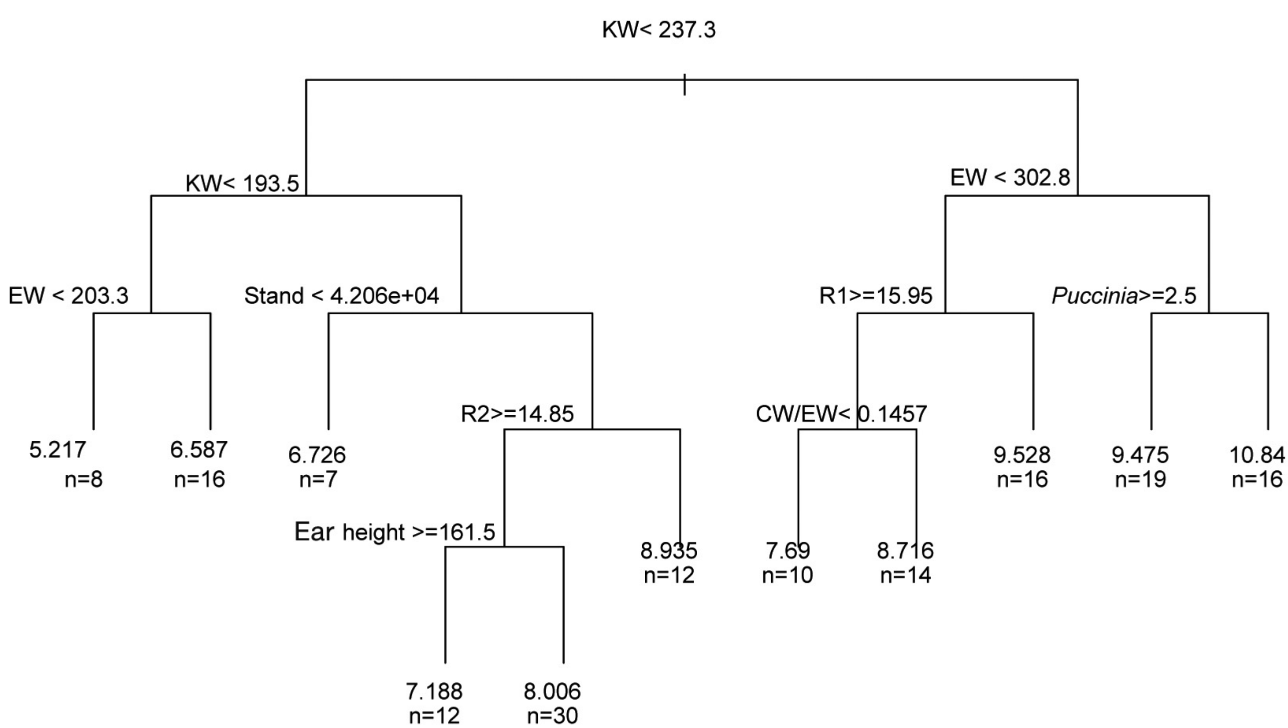

FIGURE 6 - Decision tree for the independent variable Yield.

by the linear regression model, providing significant estimates of response to selection for yield (R^2 = 83.9%).

Significant differences were found among cycles of selection for yield (cycle x environment interaction). Significant differences were found among environments (field locations) for grain moisture (Table 4).

MARS *analysis* showed no variation across cycles of selection for root and stalk lodging. (Tables 3).

Mass selection at Portugal - According to MARS analysis, cycle 5 (end of breeder selection) or cycle 11 (farmer selection) are the borderline of selection procedures for breeder and farmer. Except for slight increase in kernels per row and decrease in cob diameter 3, in all the other traits no variation across selection cycles was observed for breeder selection, contrary to the generality of traits for farmer selection.

For yield, significant changes were not observed during selection when all locations were considered. For Lousada and during the first 5 cycles (breeder selection A-B-C), however a higher tendency for response to selection existed (3.09% of gain per cycle per year) for breeder selection compared with farmer selection (0.63%, of gain per cycle per year) (Figs. 8 and 9). The differences of yield gain per cycle per year between breeder and farmer selection can be related with the choice of high moisture ears selected by the farmer compared with breeder selection. Hence, the main goal of the farmer was to maximize the ear weight, but this trait explains less than 46.7% of yield variation when random forests are used. Contrary to breeder selection, farmer selection contributed to increased grain moisture (MARS, R^2 = 8.9%) during selection for greater grain yield. This fact was highly significant at Lousada (R^2 = 80.5%; 0.62% of gain per cycle per year) (Tables 3-6).

According to MARS, the beginning and end of anthesis and end of silking increased after cycle 5, i.e., during farmer selection (R^2 = 25.2; 24.8; and 32.7%; respectively). The variation is also explained by the linear regression model (R^2 = 78.2%), where significant increase of end of days-to-silk was observed (0.41% gain per cycle per year). ANOVA showed significant differences among cycles, among environments, and for year and interactions (Table 3 and 4).

The overlapping index decreased from cycle 5 to cycle 15 and after that an increase was observed, but the coefficient of determination was very low (R^2 = 6.7%) (Table 3). For Lousada a decrease tendency was observed (R^2 = 61.3%) on the rate of 3.13% per cycle per year, which means a potential increase of allogamy (Table 6).

MARS revealed a constant and low coefficient of determination for uniformity, leaf angle, tassel branching, ear placement, root and stalk lodging and presence of diseases (*Ustilago maydis* and *Puccinia spp.*). However, plant height and ear weight, significantly increased with cycles of selection (linear regression model, R^2 = 87.2; 92.3% respectively). The ANOVA for plant and ear heights showed significant differences among environments, among years and interactions with cycles of selection. Significant differences were also detected at cycle level for ear height. In the case of Lousada, regression *analysis* showed significant increases for plant and ear heights (R^2 = 52.7; 95.3%, respectively), but this increase was more obvious for farmer selection (after cycle 5).

Ear length decreased after cycle 11, especially under farmer selection (MARS, R^2 = 20.9%). Linear Regression *analysis* for Lousada indicated also that ear length was reduced from breeder to farmer selection. A positive increase was observed for ear diameter 1 (MARS, R^2 = 70.9%) from cycle 5 to 19 and then decreased. The same was observed for kernel-row-number 1 and 2 (MARS, R^2 = 85.8 and 84.9% respectively). The linear regression *analysis* showed significant increases for ear diameters 2, 3 and 4 with a percentage gain per cycle per year of 0.51, 0.85 and 0.55% respectively (Linear regression model, R^2 = 94.9; 88.3; and 89.2%, respectively). Similar outcomes were observed for Lousada emphasizing the increase of ear diameter and row numbers 1 and 2 in the farmers' selection (Tables 3-6).

The fasciation increased from cycle 11 to 19 and then decreased (MARS, R^2 = 49.0%). At Lousada, fasciation significantly increased (R^2 = 78.8%) with 4.7% of gain per cycle per year. This is especially interesting if we consider that farmers, during seed selection, balance the choice of fasciated ears with other ears, but with a gain in ear diameter and kernel row number. The convulsion increased after cycle 11 (MARS, R^2 = 37.8%) for farmer selection. For Lousada (MARS, R^2 = 34.3%) this tendency was higher. This increase, according to GALINAT (1980), is associated with fasciation. No significant differences were observed for kernel type and ear and kernel weight. Kernel depth increased after cycle 11 (R^2 = 11.4%) under farmer selection, which can be related

TABLE 6 - *After positive assumption of normality, linear regression was used. Estimation of linear regression coefficient (b), their standard errors, initial cycle prediction ($\hat{C}0$), coefficients of determination (R^2) and % of gain per year (%Gain/Y) for mass selection (22 cycles in Portugal). For Lousada, Portugal 46 traits were collected during 2005, 2007 and 2008. Mean traits for standard populations are also included.*

	Mass selection Pt - Lousada							Populations Standard Pt -Lousada			
Traits	b	$\hat{C}0$	R^2	%C/Y	C	Y	CxY	Nutica C077	BS21(R)C9	BS22(R)C9	TEPREC6
Yield, Mg ha-1	-0.03 ± 0.03	8.76	0.17	-8.52		**	**	10.19	8.35	8.34	9.22
Grain moisture %	0.20 ± 0.04 **	31.62	0.81	0.62	**	*		32.98	28.49	26.14	26.24
Overlap Index	-0.02 ± 0.01	0.58	0.61	-3.13				0.80	0.39	0.45	0.62
Plant height, cm	1.77 ± 0.49 *	283.18	0.53	0.63	**	**	**	284.25	239.37	228.23	220.37
Ear height, cm	2.13 ± 0.21 **	150.14	0.95	1.42	**	**	**	165.43	115.13	108.43	111.60
Ear length, cm	-0.11 ± 0.03 **	22.43	0.78	-0.51	**	**	**	21.51	15.58	18.37	16.01
Ear diameter 3, cm	0.04 ± 0.01 **	4.54	0.89	0.84	**	**	**	4.77	4.42	4.23	4.29
Ear diameter 2, cm	0.03 ± 0.00 **	4.93	0.94	0.52	**	**	**	5.22	4.76	4.83	4.61
Ear diameter 4, cm	0.02 ± 0.00 **	4.51	0.87	0.55	**	**	**	4.64	4.32	4.16	4.20
Fasciation	0.06 ± 0.01 **	1.31	0.79	4.66	**	**	**	1.91	1.17	1.03	1.02
Flint/Dent	0.00 ± 0.01	6.35	0.00	-0.02	**	**		6.55	6.97	6.47	6.63
Ear weight, g	-0.36 ± 0.69	275.32	0.05	-0.13	**	**	**	292.28	177.61	206.28	177.49
Kernel weight, g	-0.33 ± 0.61	231.90	0.06	-0.14	**	**	**	247.82	154.17	176.70	152.88
Cob/Ear weigth	0.00 ± 0.00	0.16	0.00	0.01	**	**	*	0.15	0.13	0.15	0.14
Kernel dept, cm	0.00 ± 0.00	1.21	0.56	0.17	**	**	**	1.26	1.22	1.13	1.20
Kernel number, nº	4.56 ± 1.67 *	583.15	0.60	0.78	**	**	**	659.77	445.63	508.88	482.73
Thousand kernel weight, g	-2.99 ± 0.41 **	397.08	0.91	-0.75	**	**	**	376.92	350.46	347.89	319.32
Kernel per row, nº	-0.19 ± 0.04 **	41.57	0.79	-0.45	**	**	**	42.09	30.25	35.40	33.40
Cob diameter 1, cm	0.03 ± 0.00 **	3.33	0.96	0.91	**	**	**	3.56	3.02	3.19	2.85
Cob diameter 3, cm	0.03 ± 0.00 **	2.68	0.91	1.13	**	**	**	2.77	2.43	2.52	2.31
Cob diameter 2, cm	0.02 ± 0.00 **	3.15	0.98	0.52	**	**	**	3.28	2.87	3.08	2.70
Cob diameter 4, cm	0.01 ± 0.00 **	2.66	0.83	0.55	**		**	2.61	2.33	2.46	2.26
Medulla 1, cm	0.03 ± 0.00 **	1.22	0.97	2.49	**	**	**	1.41	0.87	1.20	0.98
Medulla 2, cm	0.02 ± 0.00 **	1.02	0.83	1.64	**	**	**	1.15	0.75	1.02	0.86
Raquis 1, cm	0.03 ± 0.00 **	2.38	0.99	1.18	**		**	2.43	2.00	2.23	2.03
Raquis 2, cm	0.01 ± 0.00 **	2.15	0.78	0.63	**	*	**	2.15	1.82	2.05	1.85

* - Significant at 0.05 probability levels; ** - Highly significant at 0.01 probability levels; † Number of days from date of planting to date of flowering; ‡ - the stand correspond to the average of the correspondent cycles.

%Gain/Y – percentage of gain per year, ANOVA for C-cycles of selection, E-environment; Y-years; x-interactions; $\hat{C}$ - predicted cycle of selection, except for stand that was calculated the average. Flowering data was not measured in Lousada, Portugal in 2008.

with increased fasciation. Kernel number significantly increased with selection and registered a gain per cycle per year of 0.73% (R^2 = 70.5%).

Thousand kernel weight, however, significantly decreased (R^2 = 72.7%) at a rate of -0.51% cycle/year. For Lousada this decrease was greater (R^2 = 91.4%) at a rate of -0.75% cycle/year. Hence for breeder selection there was a tendency for kernel weight to increase. The decrease of kernel weight under farmer selection is related not only with fasciation increase but also with the greater importance of one particular trait in the formula used for "Best Ear of Sousa Valley". The formula, conceived by Pego, is supposed to give the Ear Value (EV). EV is based on the kernel weight at 15% moisture (KW), ear length (L), kernel row number (R) and number of kernels (KN) [EV = (0.6 KW + 0.2 L + 0.15 R + 0.05 KN)/4].

Kernels per row showed an increase until cycle 11 and then a decrease (MARS, R^2 = 15.5%). At Lousada a significant decrease was observed (R^2 = 78.7%) with a -0.45% decrease per cycle per year.

Cob diameters 1, 2, and 4 and rachis 1 and 2 significantly increased during selection (R^2 = 96.3; 97.7; 87.7; 96.9; 76.9%, respectively). For cob diameter 3 the MARS *analysis* indicated a decrease until

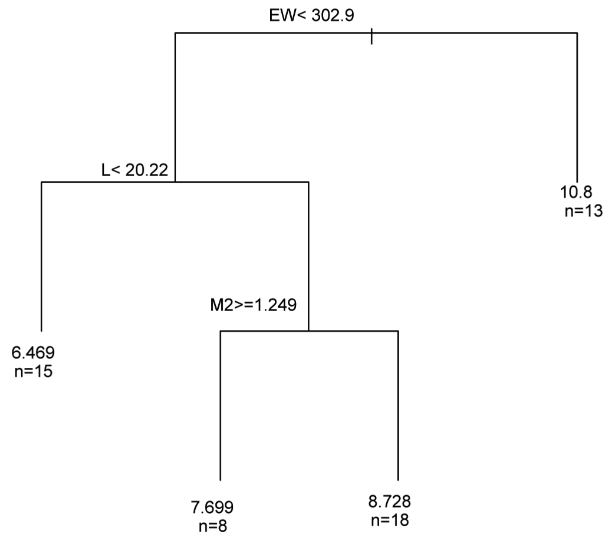

FIGURE 7 - Decision tree for the independent variable Yield for Lousada.

cycle 11, increase from cycle 11 to 19, and after cycle 22 a slight decrease. The medulla 1 increased with farmer selection (after cycle 5). At Lousada, significant increases of cob diameters 1, 3, 2 and 4, medulla 1 and 2 and rachis 1and 2 did occur and gains per cycle ranged from 0.52 to 2.49%. During selection, therefore, cobs became larger as reflected in the changes for medulla and rachis (Tables 3-6).

To better understand the causes that explain yield in 'Fandango', complementary *analysis* were based on MARS, RF and CART. The MARS approach (R^2 = 75.1%) indicated ear weight and kernel row number 1, were the most important traits to explain grain yield. The random forest approach explained 46.7% of grain yield for the variables used. Variables such as kernel and ear weight, number of kernels per row, ear length, row number 1 and 2, ear diameter 1 and thousand kernel weight were the highest ranked traits when Mean Decrease Accuracy (% IncMSE) was used. For Mean Decrease MSE (IncNodePurity) the most important variables were ear and kernel weight, ear length, number of kernels per row, stand, thousand kernel weight, number of kernels per ear, and ear diameter 2. The CART *analysis* revealed that kernel and ear weight, stand, number of kernel rows 1 and 2, *Puccinia spp*., ratio cob and ear weight and plant height were the most important traits to explain yield. Both the MARS and CART *analysis* included ear weight and kernel-row-number as important traits for grain yield.

The results using the R-project (R Development Core Team, 2008) obtained for each one of the methods are presented in Table 7 and Figs. 4 to 6.

The MARS results for Lousada (R^2 = 82.7%) showed that kernel weight, grain moisture, leaf angle insertion, and ear placement, as important traits for grain yield. The random forest approach for Lousada explained 54.4% of the yield variation in which ear and kernel weights, ear length, grain moisture, number of kernels per row, thousand kernel weight, plant height, stand and ear height, were the highest ranked traits when Mean Decrease Accuracy (% IncMSE) was used. For Mean Decrease MSE (IncNodePurity) the most important variables were ear and kernel weight, ear length, number of kernels per row, grain moisture, kernel number, stand and thousand kernel weight. The CART *analysis* revealed that ear weight and length as well as medulla 2 were used for Lousada (Fig. 7).

Standard North American populations

The standard populations showed no significant differences between Iowa and Portugal, which did not happen with 'NUTICA' and 'Fandango' cycles presenting a yield variation of -40.7 and -38.5% respectively, between Iowa and Portugal (Table 3 and Table 6). These results can be caused not only by the lack of adaptation of 'NUTICA' and 'Fandango' to Iowa environments, but also to mechanical harvest used in Iowa (high root and stalk lodging) (Table 3).

DISCUSSION

Trials in Iowa revealed a significant decrease of yield along cycles of selection, indicating that selection done at Lousada did not match with Iowa environment, considering different harvest procedures; hand in Portugal versus mechanical at Iowa. These results also indicate that during the selection process the ability of adaptation to Iowa decreased (Tables 3-4).

Response to mass selection in Portugal, revealed significant increase for silking end (R^2 = 78.21). According to MARS analysis, data related with flowering and grain moisture content increased after cycle 5, i.e., during farmer selection. Hallauer and Miranda (1988) reported that during mass selection there was a decrease of earliness that has a positive relationship with yield. The plant and ear heights increased significantly, but low correlations of heights with

TABLE 7 - *Using MARS to explain the variable Yield considering all the locations and Lousada.*

	Total	Lousada
R^2:	**75.1%**	**82.7%**
Yield=	7.905583	8.98437097
	+0.01993999*max{0,(EW-224.725)}	+0.05613638*max{0,(KW-232.736)}
	-0.02497563*max{0,(224.725-EW)}	-0.02452403*max{0,(232.736-KW)}
	+1.091053*max{0,(17.7-R1)}	-0.29788986*max{0,(MOIST-32,1)}
	-0.0001935974*max{0,(42708-Stand)}	-1.81219238*max{0,(N-5)}
	-12.22282*max{0,(CC-1.85)}	-1.32918143*max{0,(5-E)}
		-0.1129941*max{0,(MOIST-22.8)}
		+2.446074*max{0,(2-Puccinia)}
		-3.853924*max{0,(5.715-DE1)}
		-0.0134494*max{0,(SW-328.499)}
		-0.5679192*max{0,(6-N)}
		+2.978107*max{0,(DE4-4.515)}
		+3.87864*max{0,(1.35-Fa)}
		-4.968153*max{0,(DE3-4.94)}
		+7.738557*max{0,(DE4-5.095)}
		+2.688185*max{0,(M1-1.4075)}
		+0.3699823*max{0,(U-2)}

grain yield usually occur (HALLAUER and MIRANDA, 1988). The tassel size increased after cycle 11, which seems to be related with ear fasciation increase; i.e., greater size of tassel is related to fasciated ears (ANDERSON, 1944). Data related with the ear traits reveal by linear regression a significant increase of ear diameters 2, 3 and 4, kernel number, cob diameters and rachis, as in for thousand kernel weight, a significant decrease on linear regression was observed. The regression *analysis* data and MARS approach, indicates that ear evolution occurred specially under farmer selection and that these changes were mainly significant increases of ear and cob diameters and rachis. There was a tendency, according to MARS analysis, to a decrease in ear length and increases of kernel-row-number, convulsion and fasciation expression, which agrees with reports by HALLAUER and MIRANDA (1988) and PEGO (1982).

For Lousada, the location where breeding was done, the fasciation trait and medulla size significantly increased with selection, whereas ear length and kernels per row significantly decreased. Similar outcomes were observed in long-term divergent selection for ear length in maize (HALLAUER, 1992) and by EMERSON and EAST (1913) for relations between ear length and number of kernel-rows and between ear diameter and kernel-rows number and seed size. The kernel row arrangement became significantly more irregular (convulsion), which could be related with fasciation (Tables 5-6).

The selection process included 22 phenotypic mass selection cycles and occurred in two phases: 1) The breeder phase from cycle 1 to cycle 5, and 2) The farmer phase, after cycle 5.

The aim of the breeder was the yield improvement of 'Fandango'. To achieve this goal, stratified mass selection was done for both parents. For yield, no significant changes were observed during selection when all locations were considered (Fig. 8). Nevertheless for Lousada, and during the first 5 cycles, a higher tendency exists for yield increase (3.09% of gain per cycle per year) for breeder selection compared with farmer selection (0.63%, of gain per cycle per year) (Fig. 9).

The aim of the farmer selection was the ear size maximization. This selection procedure can be related to: a) hand versus mechanical harvesting. Generally farmers prefer lower densities and bigger ears if they harvest by hand; b) the "Best Ear of the Sousa Valley competition", was one of the main reasons that explains the popularity of 'Fandango'. Hence during farmer selection some decisions could prejudice hypothetical yield gain, such as the selection of higher moisture ears (for Lousada, R^2 = 80.5; 0.62% of gain per cycle per year) comparing with breeder selection. Considering that maximum ear size is

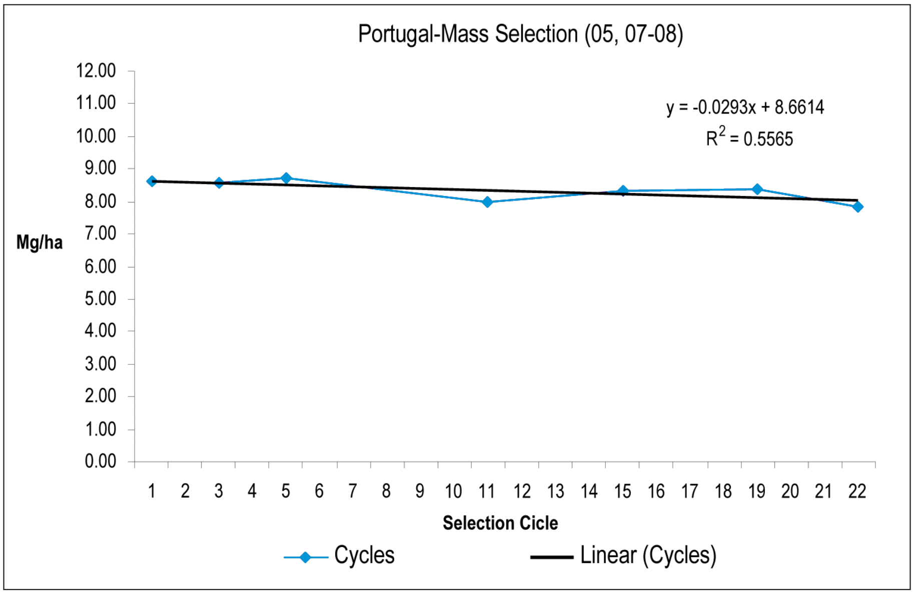

FIGURE 8 - Yield evolution during the 22 cycles of mass selection.

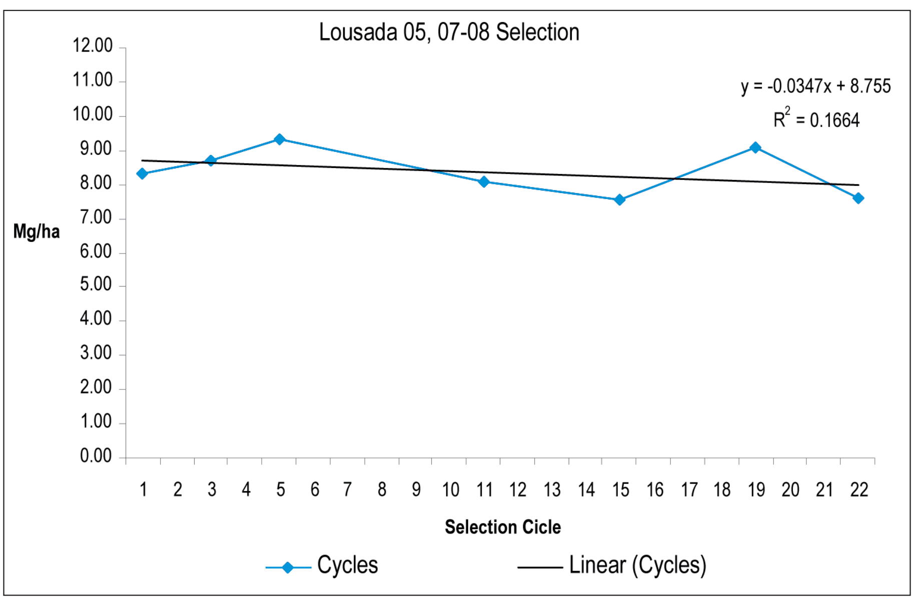

FIGURE 9 - Yield evolution during the 22 cycles of mass selection for Lousada. The first five cycles represent the breeder selection.

highly related with ear weight, this trait for 'Fandango' explains less than 46.0% of yield variation when random forests are used. 'Fandango' is not adapted to high densities. During selection plant and ear height significantly increased, which could mean less area available, *i.e.*, competition in trials was more severe to advanced cycles and some plants did not produce ears. Probably for this reason significant decrease in yield was observed at Iowa locations. In general the lack of significant progress in yield for phenotypic mass selection could be also explained by the low selection intensity due to the exclusion of stalk lodged plants in the basic units of selection. Hallauer and Sears (1969) observed that in the absence of a correlation between grain yield and stalk lodging, the exclusion of stalk lodged plants reduces the intensity of selection for yield from 7.5 to 27.4%.

Despite the absence of significant yield progress, mass selection in Portugal increased significantly the number of days to silk, plant and ear heights, and ear size (significant increase for ear diameter, kernel number, cob and rachis diameters) and decreased significantly the thousand kernels weight. For Lousada, fasciation and medulla also increased significantly and ear length and kernels per row decreased significantly. Identical outcomes were observed in long-term divergent selection for ear length in maize (Hallauer, 1992).

Thousand kernels weight significantly decreased with cycles of selection, but for the breeder selection there was a tendency for thousand kernels weight to increase. The generalise decrease of thousand kernels weight could be related, not only because of fasciation pressure, but also for the importance of number of kernels per ear in the formula used for "Best ear of Sousa Valley" by farmers.

The fasciation evaluation suggests that the farmer emphasize fasciation during selection to increase ear diameter and kernel row number. Level of ear fasciation is especially interesting at Lousada (R^2 = 78.8%) with 4.7% increased fasciation per cycle/year. During seed selection, farmers keep fasciated ears in certain proportion to make a bulk with certain equilibrium of level of ear fasciation expression.

RF, CART, and MARS *analysis* revealed that kernel weight and ear weight were the most important traits for grain yield expression, but row numbers, number of kernels per row, ear length, and ear diameter were also some of the important traits that influence 'Fandango' yield. The proper balance of these six components for grain yield expression will be attained by greater precision in selection of ears having the greatest yield.

FUTURE PERSPECTIVES

The lack of significant progress in grain yield for 'Fandango' suggests new experiments for the future should be pursued: greater parental control of plants included in selection, plant density trials either in monocrop or in polycrop systems, fertilization level trials, extension of the studies of overlapping index (Moreira and Pego, 2003). Hybrid populations' development could contribute also to yield progress and to avoid the collapse of some interesting germplasm. Its link with a PPB program offers also an opportunity to better design synthetic hybrid populations for low input and organic agriculture.

Molecular data input will be added in the future to clarify: 1) what happened to 'NUTICA' during recombination and selection (using the original inbreds until the formation of 'NUTICA'); 2) the understanding of the evolutionary process from 'NUTICA' to 'Fandango'; and 3) the evolution of the genetic diversity of 'Fandango' during breeder selection (cycle 5) and farmer selection. These studies could help also to find the possible existence of association between particular molecular markers and some of the phenotypic traits under study (e.g., ear length, ear diameter, kernel-row number and fasciation). The identification of molecular markers suited for marker assisted selection would be useful, but more research is needed. Also the genetic control of some of the phenotypic traits here evaluated (such as the fasciation trait) is under study.

Besides being an interesting population for farmers, 'Fandango' is intrinsically linked with the contest of "Best Ear of Sousa Valley Region", because, since its beginning, 'Fandango' as been a consistent winner in the yellow dent group. This competition is a powerful tool for breeder as a:

1) Pedagogic tool: throughout the ear value formula the breeder can indirectly indicate to the farmer what are the most important traits and their relative importance for selection in their own populations, e.g. kernel weight, ear length, kernel row number and total number of kernels. While kernel depth is also an important parameter related with yield, it is supposed to be indirectly covered by the four parameters included in the formula.
2) Germplasm "tracker": during farmers' inscription for competition information data is registered, which allows the breeder to find the farmer in order to obtain a sample of his germplasm and valuable data (e.g., 'Verdeal de Aperrela' was included in VASO project throughout this method), that could be used to evaluate the level of rural development and level of desertification.
3) Germplasm "disseminator": after competition ears remain in the cooperative of Paredes, which provides an effective method of dissemination.
4) Social aspects: this contest permits the recognition of the farmer by the community, but also attract new farmers and germplasm for new initiatives.

Compared with the literature on collaborative plant breeding, VASO can be considered exemplary in regards to its duration. But similar to other areas, this project recognizes that the future of smallholder farming as a viable way of life in Portugal is decreasing due to the socio-economic "pull" factors that remove younger generations from the farm (Powell, 2000; Vaz Patto *et al.*, 2007).

Considering the definition of maize breeding by Hallauer and Carena (2009), 'Fandango' as a fasciated population is really "the art and science of compromise". The farmers and specially Mr. Meireles were able to be artists for developing greater size ears by emphasizing the ear fasciation trait which is a difficult trait to use in selection.

ACKNOWLEDGEMENTS - Particular acknowledgements are due to the farmer Mr. Meireles and to Dr. Wayne Haag (as CIMMYT director of maize breeding in the Mediterranean area in 1985) who first supported the VASO project. We also acknowledge Engº Vaz Patto, Engª Rosa Guilherme from ESAC, Engº Mário Pardal from Escola Profissional Agrícola Afonso Duarte and Prof. Kendall Lamkey and team from Iowa State University, for his kind support with trials and data collection in Iowa and to Mary Lents. To Nuno Sousa. To Daniela, Lucas and Sara. This research was supported by long-term financial support by Câmara Municipal de Lousada and by the following projects of Fundação para a Ciência e a Tecnologia, Portugal: POCI/AGR/57994/2004, PTDC/AGR-AAM/70845/2006; P. Mendes Moreira was supported by SFRH/BD/22188/2005.

REFERENCES

ANDERSON E., 1944 Homologies of the ear and tassel in *Zea mays*. Ann. Missouri Bot. Garden **31:** 325-343.

BREIMAN L., 1996 Bagging predictors. Machine Learning **26:** 123-140.

BREIMAN L., 2001 Random forests. Machine Learning **45:** 5-32.

BREIMAN L., J.H. FRIEDMAN, R.A. OLSHEN, C.J. STONE, 1984 Classification and regression trees. Chapman and Hall/CRC.

EMERSON R.A., E.M. EAST, 1913 The inheritance of quantitative characters in maize. Univ. Nebraska Res. Bull. Nº. 2 (cit. in Pego, 1982).

FERRÃO J.E.M., 1992 A aventura das plantas e os descobrimentos Portugueses. Programa Nacional de Edições Comemorativas dos Descobrimentos Portugueses.

FRIEDMAN J.H., 1991 Multivariate adaptive regression splines. Annals of Statistics **19:** 1-141.

GALINAT W., 1980 Indeterminated *vs* determinated years. Maize Genet. Coop. Newsletter **54:** 121.

HALLAUER A.R., 1992 Recurrent selection in maize. pp. 115-177. *In:* J. Janick (Ed.), Plant Breeding Reviews. Vol 9. J. Wiley and Sons, Inc.

HALLAUER A.R., 1994 Corn genetics and breeding. Encyclopaedia Agric. Sci. **1:** 455-467.

HALLAUER A.R., J.H. SEARS, 1969 Mass selection for yield in two varieties of maize. Crop Sci. **9:** 47-50.

HALLAUER A.R., J.B. MIRANDA F.O., 1988 Quantitative genetics in maize breeding. 2nd edn. Iowa State Univ Press, Ames.

HALLAUER A.R., M. CARENA, 2009 Maize breeding. *In:* M.J. Carena (Ed.), Cereals. Springer.

HALLAUER A.R., W.A. RUSSELL, P.R. WHITE, 2000 Registration of BS21(R)C6 and BS22(R)C6 maize germplasm. Crop Sci. **40:** 1517.

HALLAUER A.R., A.J. ROSS, M. LEE, 2004 Long-term divergent selection for ear length in maize. pp. 153-168. *In:* J. Janick (Ed.), Plant Breeding Reviews. Vol 24. J. Wiley and Sons, Inc.

HASTIE T., R. TIBSHIRANI, J.H. FRIEDMAN, 2001 The elements of statistical learning: data mining, inference, and prediction. Springer.

KUHN J., B. EGERT, S. NEUMANN, C. STEINBECK, 2008 Building blocks for automated elucidation of metabolites: Machine learning methods for NMR prediction. BMC Bioinformatics **9:** 400.

LILLIEFORS H.W., 1967 On the kolmogorov-smirnov test for normality with mean and variance unknown. J. Am. Statistical Ass. **62:** 399-402.

LONNQUIST J.H., 1961 Progress for recurrent selection procedures for improvement of corn populations. Nebrasca Agric. Exp. Stn. Res. Bull. 197.

MENDES MOREIRA P.M., M.C. VAZ PATTO, S.E. PEGO, A.R. HALLAUER, 2008 Comparison of selection methods on 'Pigarro', a Portuguese improved maize population with fasciation expression. Euphytica DOI:10.1007/s10681-008-9683-8.

MOREIRA P.M., 2006 Participatory maize breeding in Portugal. A case study. Acta Agron. Hungarica **54:** 431-439.

MOREIRA P.M., S. PEGO, 2003 Pre-breeding evaluation of maize germplasm. The case of a Portuguese open-pollinated variety. *In:* Abstracts of the A.R. Hallauer Intl. Symposium on Plant Breeding. Mexico City, Mexico 17-22 August 2003.

MOTA M., E. BETTENCOURT, L. GUSMÃO, 1978 Cruzamentos múltiplo em milho. pp. 134-135. *In:* Relatório das Actividades, Estação Agronómica Nacional.

PEGO S.E., 1982 Genetic Potential of Portuguese maize germplasm with abnormal ear shape. Ph.D. Dissertation, Iowa State University.

PEGO S.E., A.R. HALLAUER, 1984 Portuguese maize germplasm with abnormal ear shape. Maydica **29:** 39-53.

PEGO S.E., M.P. ANTUNES, 1997 Resistance or tolerance? Philosophy, may be the answer. *In:* Proc. XIX Conference Intl. Working Group on Ostrinia. Guimarães Portugal 30th August-5th September 1997.

POWELL J., 2000 The relationship between ideotypes, knowledges and practices in plant breeding among farmers and scientists. Society for Social Studies of Science (4S) and European Association for the Study of Science and Technology (EASST), Vienna, Austria.

R DEVELOPMENT CORE TEAM, 2008 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

TILMAN D., K.G. CASSMAN, P.A. MATSON, R. NAYLOR, S. POLASKY, 2002 Agricultural sustainability and intensive production practices. Nature **418:** 671-676.

TROYER A.F., 2000 Origins of modern corn hybrids. Proc. Ann. Corn Sorghum Res. Conf. **55:** 27-42.

VAZ PATTO M.C., P.M. MOREIRA, V. CARVALHO, S.E. PEGO, 2007 Collecting maize (*Zea mays* L. convar. mays) with potential technological ability for bread making in Portugal. Gen. Res. Crop Evol. **54:** 1555-1563.

WOLFE M.S., P. BARESEL, D. DESCLAUX, I. GOLDRINGER, S. HOAD, G. KOVACS, F. LÖSCHENBERGER, T. MIEDANER, H. ØSTERGÅRD, E.T. LAMMERTS VAN BUEREN, 2008 Developments in breeding cereals for organic agriculture. Euphytica **163:** 323-346.