Effectiveness of Selection Methods for Improvement of Portuguese Maize (*Zea mays* L.) The study of fasciation

Pedro Manuel Reis Mendes Moreira

Dissertation presented to obtain the Ph.D degree in Biology (Population Genetics)

Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras, "July 2015"

INSTITUTO DE TECNOLOGIA QUÍMICA E BIOLÓGICA /UNL

Knowledge Creation

PhD Supervisors:

Dr. Maria Carlota Vaz Patto (Supervisor)

Prof. Pedro Fevereiro (Co-Supervisor)

The work presented in this thesis was performed at:

Instituto de Tecnologia Química e Biológica António Xavier (ITQB)

Instituto Politécnico de Coimbra - Escola Superior Agrária de Coimbra

PhD thesis Advisory Board

Prof. Arnel R.Hallauer

Dr. Silas E. Pêgo

The student, Pedro Manuel Reis Mendes Moreira received financial support from Fundação para a Ciência e Tecnologia and Fundo Social Europeu in the scope of Quadro Comunitário de Apoio through the PhD fellowship SFRH/BD/22188/2005

and

in the scope of programa de apoio à formação avançada de docentes do Ensino Superior Politécnico SFRH/PROTEC/50365/2009

Additional Funding:

FCT project « Analysis of the breeding potential of fasciation traits in Portuguese maize landraces » (PTDC/AGR-AAM/70845/2006) (2007-2010)

Chapter III presentation was sponsored by Cost Action 860

Short-term grant by Zagreb University, under guidance of Zlatko Šatović

ERASMUS Program
OpenDipSA - ERASMUS 22-23 July 2013
University of Bologna

European Commission through the FP7-KBBE-2009-3, under the grant agreement no. 245058, SOLIBAM project

“Advances in medicine and agriculture have saved vastly more lives than have been lost in all the wars in history.”

— Carl Sagan, The Demon-Haunted World: Science as a Candle in the Dark

To Daniela, Lucas, Sara and Francisca

Acknowledgments

After my stage at NUMI/BPGV since 1995 to 1997 under the guidance of Silas Pêgo, many attempts were made to start the PhD program. In 29 March 2003 the start of this journey was initiated with the submission of a poster to Arnel R. Hallauer International Symposium on Plant Breeding held at the Hotel Camino Real in Mexico City from 17-22 August, 2003. At this Symposium I had the opportunity to meet Arnel R. Hallauer, which resulted in two actions, my stage in 2004 at Iowa State University at Kendall Lamkey maize breeding program with a FLAD support and in the invitation to Arnel R. Hallauer to come to Portugal that was held at same year. During this time I had the opportunity to invite Arnel R. Hallauer as PhD. Advisor together with Silas E. Pêgo. In parallel the project POCTI/AGR/57994/2004 "Portuguese maize bread quality: from technological aspects to controlling genes", had started in 1st of February of 2005 with Maria Carlota Vaz Patto as Project Coordinator. Due to Maria Carlota Vaz Patto expertise and interest on the maize breeding topic she was the skilled and right person as PhD Supervisor together with Pedro Fevereiro as PhD Co-supervisor on the molecular analysis that were missing on the fasciation trait. This was also the reason that we apply to both applications to FCT, both succeeded, throughout PhD fellowship and later by the project «Analysis of the breeding potential of fasciation traits in Portuguese maize landraces»

(PTDC/AGR-AAM/70845/2006) coordinated also by MC Vaz Patto, that were crucial for the development of this thesis.

This story gives the context to acknowledge to my Supervisors and Advisors for their acceptance and valuable contributions, which were crucial in each step of PhD work. Without them this work would not be possible.

- Silas Pêgo, for his constant help motivation and availability, guidance, valuable corrections and suggestions and wisdom for starting the **VASO PPB Project**, one of the oldest in the world.
- Arnel Hallauer, for all the support, guidance and valuable corrections and suggestions of the manuscripts.
- Pedro Salema Fevereiro, for his acceptance and support in the PhD work.
- Carlota Vaz Patto for her constant help, support and guidance, since the lab work interpretation of results and a crucial help on the paper submissions and thesis elaboration.

My experience at Iowa State University with Arnel Hallauer is naturally extended to Kendal Lamkey and team from Iowa State University that kindly accepted me in his program and later he kindly accepted to perform part of multilocation trials in USA during 2005, described in Chapter III and Chapter IV.

To BCV, ITQB personal (Nuno Almeida, Mara Alves, Rita Caré) for their untiring tutoring, support and teachings in the lab, constant encouragement and meaningful advice before and after the lab.

To Zlatko Šatović, for his kind support, and fruitfull discutions at Zagreb University via his guidance and scholarship that made possible and very pleasant my stage at his lab, 谢谢.

To João Pedro Mendes Moreira for the presential and tele meetings that were very important to create a common language of discussion especially important for Chapter VI.

To my students and collaborators: João Santos, J.P. Santos, Marisa Antunes, Liliana Morgado, Ralf Moura, Cláudia Brites, Manuel Paulo, Daniel Gaspar, Graciano Spencer, Joana Laires …

To collaborators Engº Vaz Patto, Engª Rosa Guilherme from IPC-ESAC.

In memoriam to Wayne Haag (as CIMMYT director of maize breeding in the Mediterranean area in 1985) who first supported the VASO project.

To the farmers and specially Mr. Francisco Meireles for its critical analysis and empirical knowledge, that made VASO project a bridge between empirical knowledge and science… and probably a "seed" for Socionomics in Participatory Plant Breeding.

To my friends that do not see me for a while … even the nearest ones… but that were encouraging and supporting me all the time.

To my parents for their solid contribution to my G, E and G x E. To my father for his constant learning encouragement and perseverance.

To Daniela, Lucas, Sara and Francisca that also started the PhD with me…

A Todos o meu Muito Obrigado.

Sumário

Desde a sua introdução, pós Colombo, o milho provocou uma revolução silenciosa sobre a região Norte e Centro de Portugal. Revolução que se traduziu pela reformulação dos sistemas de cultivo, agronomia, paisagem e cultura ao longo dos anos.

Na década de 1940, o sucesso das sementes híbridas americanas iniciou o seu contributo para a erosão genética. No NUMI (Estação de melhoramento de milho em Braga) Silas Pêgo compreendeu desde logo esta ameaça. Assim, várias missões de colheita de germoplasma de milho foram organizadas. Esta recolha contribuiu para a conservação *ex-situ*; tendo contribuído também para as atividades *in situ* / on-farm e on-station *via* pré-melhoramento.

O presente trabalho inicia-se com a descrição do projeto VASO (um projeto de Melhoramento Participativo de Plantas), iniciado em 1984 na região do Vale do Sousa onde a conservação da diversidade genética e atividades de melhoramento continuam, tendo por objetivo a qualidade do milho para "broa" (pão de milho). O projeto VASO representa também uma oportunidade para a adaptação do germoplasma às áreas marginais de produção, agricultura sustentável e integração do conhecimento tradicional.

O nosso trabalho prosseguiu com uma caracterização detalhada do trabalho de melhoramento participativo do projeto VASO desde a sua génese. A abordagem quantitativa seguida, permitiu comparar os

métodos de seleção aplicados pelo melhorador e agricultor, utilizando 'Pigarro' "(variedade regional portuguesa de milho liso, branco) e 'Fandango' (uma população sintética de milho amarelo dentado). Como resultado, os agricultores selecionaram espigas mais curtas e largas, com aumento dos níveis de fasciação e grãos de menor dimensão. No caso da seleção do melhorador, as espigas tornaram-se mais compridas e menos fasciadas, com um aumento da uniformidade da cultura. Ambos os métodos de seleção do melhorador e agricultor foram eficazes para a conservação da diversidade. Deste modo, a escolha do método de seleção, dependerá dos objetivos do programa de melhoramento: seleção fenotípica recorrente é mais fácil e potencialmente mais económica para adotar pelos agricultores para melhoramento de OPVs (variedades de polinização livre), enquanto que os resultados de seleção pelo melhorador resultam numa maior uniformidade da cultura, estando mais adaptadas para programas de desenvolvimento de híbridos.

A análise da evolução da diversidade molecular enfatizou associações potenciais entre determinados marcadores moleculares neutros e os *loci* responsáveis pelo controlo de algumas das características fenotípicas sob seleção (*e.g.*, comprimento da espiga, fasciação e características associadas à espiga como o diâmetro da espiga e número de grãos por carreira). Estas associações precisam no entanto de ser melhor analisadas e validadas através de

mapeamento de ligação ou associação, de modo a ancorarem a seleção de características a sistemas de agricultura sustentáveis.

O nosso trabalho permitiu ainda que os dados fenotípicos fossem utilizados no desenvolvimento de ferramentas de seleção para os agricultores, contribuindo para melhorar o processo de seleção. Com este propósito, melhoramos a "fórmula do valor da espiga", como ferramenta de seleção do agricultor, procurando aumentar a produção com base nas características da espiga. O valor da Espiga (EV) foi desenvolvido em 1993 no âmbito de um concurso regional de espigas de milho (Concurso da “Melhor Espiga do Vale do Sousa”). Esta fórmula tinha dois objetivos principais, a avaliação de espigas para o referido concurso e como ferramenta pedagógica no melhoramento de milho por parte dos agricultores. A fórmula EV foi baseada em correlações de características de milho publicadas na literatura, sem que constassem dados de campo como a produção. Para cumprir esta lacuna geramos métodos diferentes e desenvolvemos um método de classificação para compará-los. Utilizamos para tal os dados de um conjunto de populações de onde as melhores espigas do Vale do Sousa provieram. A partir dos métodos utilizados, a fórmula EVA foi a escolhida por ser facilmente adotada pelos agricultores e associações interessadas na conservação e no desenvolvimento de germoplasma.

Por último, sendo a fasciação da espiga uma característica quantitativa que tem sido continuamente selecionada por agricultores portugueses, mas para a qual não existiam estudos

moleculares. O nosso objetivo centrou-se num melhor conhecimento quer a nível molecular quer fenotípico da fasciação da espiga, cuja variação morfológica, pode ter um impacto efetivo sobre a produção. Para cumprir esta lacuna utilizamos a população $F_{2:3}$, desenvolvida a partir de um cruzamento entre linhagens divergentes (PB260 não fasciada x PB266 fasciada) por forma a elucidar as características genéticas da fasciação da espiga. Foi detetada variação significativa entre linhagens parentais PB260 e PB266 e foram mapeamos uma série de QTLs que controlam características relacionadas com a fasciação.

O QTL constitutivo detetado para fasciação localizou-se no cromossoma 7, indicando *ramosa3* (*ra3*) como um gene candidato. Além disso, este estudo de mapeamento de QTLs contribuiu para expandir a lista de áreas genômicas potencialmente envolvidas na fasciação da espiga de milho e características relacionadas, especialmente nos cromossomas 1, 3, 5, 7 e 8, onde outros genes candidatos *barren inflorescence2* (*bif2*), *ramosa2* (*ra2*), *tasselseed4* (*ts4*), *terminal ear1* (*te1*), *bearded-ear1* (*bde1*), *branched silkless1* (*bd1*) and *compact plant1* (*ct1*) foram propostos, utilizando marcadores moleculares neutros selecionados como flanqueadores. Verificou-se que algumas das associações detetados no 'Pigarro' ocorreram igualmente na população segregante PB260 x PB266 para umc1907, umc1524 e umc1858, onde *te1* e *bde1* foram considerados como genes candidatos, assim como *defective kernel19*, *28* (*dek19*, *28*), and *miniature seed3* (*mn3*).

Em resumo, o trabalho aqui apresentado apresenta: 1) uma visão mais aprofundada sobre a evolução a longo prazo do milho sob melhoramento participativo no âmbito do projeto VASO; 2) ferramentas para identificar as características fenotípicas que melhor explicam a produção através do desenvolvimento de um modelo de previsão; 3) uma compreensão adicional do papel da fasciação e genes que a controlam, a fim de expandir a lista de áreas genômicas potencialmente envolvidos na fasciação da espiga de milho e características relacionadas.

Abstract

Since its introduction after Columbus, maize provoked a silence revolution on north and central region of Portugal, reshaping crop systems, agronomy, landscape and culture along the years. In the 1940s' the advent of American hybrid seeds success started to contribute to genetic erosion. At NUMI (Maize Breeding Station at Braga) Silas Pêgo understood this threat and several maize collecting missions were organized. This collecting missions, paved the way for ex-situ conservation. In addition they feed in-situ/on farm and on station activities via prebreeding.

The present work begins with the description of the VASO project (a Participatory Plant Breeding project) initiated in 1984 at Sousa Valley Region and where genetic diversity conservation and breeding activities continues, focused maize quality for maize bread ("broa"). In addition it presents the opportunities to the adaptation to marginal areas of production, to sustainable agriculture and integrating traditional knowledge.

The work continued with a detailed characterization of the long term participatory plant breeding work at VASO project. A quantitative approach to compare the evaluation of the applied farmer's and breeder's selection methods, both using 'Pigarro' (a white flint Portuguese maize landrace) and 'Fandango' (a maize synthetic population). As a result farmers selected for shorter and wider ears, with increased levels of fasciation and smaller kernels. In the case of

breeder selection, ears became longer and less fasciated, with an overall increase of crop uniformity. Both farmer's and breeder's selection methods were effective for diversity conservation, but their choice depend on maize breeding program aims: Phenotypic recurrent selection is easier and potentially cheaper to adopt by farmers for OPV (Open Pollinated Varieties) improvement, whereas breeder selection results in a more uniform crop, being more adapted to hybrid development programs.

Our molecular diversity evolution analysis emphasized potential associations between particular neutral molecular markers and the *loci* controlling some of the phenotypic traits under selection (*e.g.*, ear length, fasciation and related ear traits as ear diameter and kernel-row number). These associations need however to be better explored and validated by future linkage or association mapping approaches previous to their use for supporting trait selection in sustainable farming systems.

Furthermore we also used phenotypic data to develop farmers' selection tools, helping farmers on selection procedures. With this purpose we improved the existent "ear value formula" as a farmer's selection tool to increase yield based on ear traits.

Ear value (EV) formula was developed in 1993 under the scope of a Portuguese regional maize ear competition (the "Sousa Valley Best Ear Competition"). This formula had two main purposes, ears evaluation for the ear competition and a pedagogical tool for maize selection for farmers. EV formula was based on published maize trait

correlations, with no direct inputs from farmers maize yield. To fulfill this gap we generate different methods and develop a ranking method to compare them using a set of populations where the best Sousa Valley ears came from. From the methods used, EVA formula was chosen, because it can be easily adopted by farmers and associations interested in germplasm conservation and development.

Lastly, being ear fasciation a quantitative trait that has been continuously selected by Portuguese farmers and for which no molecular studies existed before. It was our goal to contribute both at phenotypic and molecular level to better understand ear fasciation that despite its morphological variation, can have an effective impact on yield.

To fulfill this gap an $F_{2:3}$ population, was developed from a cross between contrasting inbred lines (non fasciated PB260 x fasciated PB266) towards the elucidation of the genetics of the fasciation trait. We have detected significant variation among parental inbred lines PB260 and PB266 and we mapped a number of QTLs controlling fasciation related traits.

The constitutive QTL detected for fasciation was located in chromosome 7, indicating *ramosa3* (*ra3*) as a putative candidate gene. In addition, this QTL mapping study has contributed to expand the list of genomic areas potentially involved in maize ear fasciation and related traits, especially in chromosomes 1, 3, 5, 7 and 8 where other candidate genes *barren inflorescence2* (*bif2*), *ramosa2* (*ra2*), *tasselseed4* (*ts4*), *terminal ear1* (*te1*), *bearded-ear1* (*bde1*), *branched*

silkless1 (*bd1*) and *compact plant1* (*ct1*) were proposed, with flanking selecting neutral molecular markers. We found that some of the associations detected for 'Pigarro' occurred also in the segregating PB260 x PB266 population for umc1907, umc1524 and umc1858, where *te1* and *bde1* were considered as candidate genes so as *defective kernel19, 28* (*dek19*, *28*), and *miniature seed3* (*mn3*).

The work here presented provides: 1) further insight into the long-term evolution of maize under participatory maize breeding project VASO; 2) tools to identify the phenotypic traits that better explain yield and the development of a prediction model for yield; 3) additional understanding of the role of the fasciation and genes that control it in order to expand the list of genomic areas potentially involved in maize ear fasciation and related traits

List of most used abbreviations

CART - Classification and Regression Trees

MARS -Multivariate Adaptive Regression Splines

NUMI - Maize Breeding Station (Núcleo de Melhoramento de Milho)

OPV - Open-pollinated Variety

PMB - participatory maize breeding

PPB – Participatory Plant Breeding

QTL Quantitative Trait *Locus*/*Loci*

RF - Random Forests

VASO - Sousa Valley, Portugal

Key words

‘Fandango’, ‘Pigarro’, fasciation, genetic diversity, integrant philosophy, maize, mass and recurrent selection, on-farm breeding, open-pollinated varieties, open-pollinated varieties, participatory plant breeding, QTL mapping, synthetic variety, VASO, *Zea mays*

Table of contents

List of figures

List of tables

CHAPTER I.

General Introduction

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
/UNL

Knowledge Creation

I.1. Portuguese maize facts

The world cereal production in 2013 was 1980 tera grams from which, 972 tera grams were maize, representing 49.1% of world cereal production (IGC 2014). Portugal, with 932 giga grams represents 0.096% of the total production in the world in an area of approximately 150 000 ha (146 719 ha in 2013 from which 101 766 ha were for grain production) (INE 2013). The new growing areas are being fed especially by the new irrigated areas (*e.g.* Alqueva). Maize in Portugal represents approximately 40% of the total area of the cereals and more than 80% of the cereals production. Portugal has been able to produce 1/3 of its needs, importing the rest 2/3, *i.e.*, Portugal has a deficit in maize. However, contrary to what happens with many other countries, Portugal has an important legacy of traditional maize varieties specifically targeted for human consumption, some still preserved on farm, others in national genebanks, but usually without a strategy for long-term use. For a country that produces around 0.1% of the world production and has genetic resources able to establish a maize breeding program and a good relationship with countries that use also maize for human use (*e.g.* CPLP) it would be useful and strategic to support participatory plant breeding and management research, development and demonstration not only in Portugal but worldwide (*e.g.* CPLP). The genetic resources legacy could differentiate us in the production of maize for human consumption as a different product compared with maize for feed. These genetic resources, initially adapted to

traditional farming systems can be adapted to different farming systems (*e.g.* organic or low input farming), being also a potential source of genes for pest and diseases and climate changes. The national plant propagation sector (seed and plants) represented in 2013, 129.4 million euros, corresponding to 1.9% of the Portuguese agricultural GDP. Additionally in 2013, 24.88 million euros of seeds and plants were exported (8.43 million for seeds), but imported 125.37 million euros (Dias 2009; INE 2014 a,b). For maize 18.52 million were imported and 2.53 million were exported. These facts should serve as a reflection on the future paths to follow, either in research and teaching, *i.e.*, how to connect the use of germplasm breeding to final products, with a rural development strategy.

I.2. Some history

The Portuguese maize germplasm introduction occurred more than five centuries ago. It is referred that maize was first cultivated in Europe on the fields of Seville and then it was introduced in the fields of Coimbra Region (Ferrão 1992).

Maize shaped Portuguese landscape (*e.g.* terraces, irrigating systems, corn cribs) and contributed to the improvement of livelihood (*e.g.* maize was available directly for human consumption as maize bread – 'Broa' – and indirectly by animal consumption). These five centuries, since maize introduction, were especially relevant to generate diversity. Diversity creation was driven by two main forces: (i) environment (Portugal has a very diverse climate, mainly due to

orography and to the Atlantic and Mediterranean influence); and (ii) farmers' selections (*e.g.* mainly plant and ear traits).

This diversity started to decrease as the American hybrids were introduced in Portugal, after the Second World War. The FAO programs for hybrid production in Europe were implemented in Portugal. The tested hybrids had excellent adaptation to Portugal and breeding stations were established along the country from north to south, but only NUMI at Braga (NUMI – maize breeding station) survived for a longer period. NUMI success was driven by its special orientation for grain quality for human use as bread and early-maturing varieties adapted to highly intense polycropping systems.

From 1982 to 1985, Silas Pêgo was responsible for the Maize National Program and, together with his mentor, Luís Costa Rodrigues, organized the National Breeding Program, with two main components: 1) On-station program, 2) On-farm program, *i.e.* 1) Monocrop System (hybrid program), adapted to the Productivist Philosophy, and 2) Polycrop System (breeding populations), adapted to the Integrant Philosophy (Pêgo, Antunes 1997). The integrant Philosophy intended to solve the problems faced by small Portuguese farmers, with both scarcity of land and highly populated areas.

In 1992 a grassroots competition "Sousa Valley Best Ear Competition" started. On the first year competition, only the number of kernels per ear was evaluated. Silas Pêgo saw the opportunity of evaluate ears in competition and tested an empirical formula in 1993 that could contribute simultaneously for: (1) farmers' maize ears traits

perception (*e.g.* What is the row number, the kernel depth and kernel weight of an ear? but also what type of kernels, for farmers, define the maize quality for maize bread?...); (2) understanding the best traits combination to select for yield improvement; and (3) providing a tool to evaluate and rank the maize ears. This provided the empirical and scientific knowledge convergence to obtain the best solution for farmer selection.

When our work started it was necessary to resume the VASO project history (one of the pioneers in participatory plant breeding project). Additionally it was also needed to evaluate VASO results, create new selection tools for farmers and define future plans.

I.3. The world context

Plant domestication is intrinsically related with the beginning of agriculture and it occurred in the world in different time frames, starting in West Asia 10 500 years ago. Its success explains the capacity to pass from 4 million when farming started to 6000 million people presently. Malthus in "An Essay on the Principle of Population in 1798" systematizes the binomial problem of food and population growth vaticinating a human population always hungry and therefore malnourished. This scape from Malthusianism was possible, because XXth century brought important discoveries in agronomy (*e.g.* Haber-Bosh process), plant breeding and genetics (*e.g.* "hybrid vigor" concept by Schull or the dwarf genes for rice and wheat) (Trewavas 2002). The adaptation of these discoveries was baptized as "Green

Revolution" and its results enabled cereal yields to increase threefold since 1950. "Green Revolution" saved many lives in the world, allowed that chronically famine countries started to be self-sufficient and stalled the expansion of new farmland needs. As example in USA, from 1944 to 2007, the total production of maize was respectively 58.42 to 332.74 Tg on approximately the same area of 34.40 million hectares. On this area a 360% increase of yield was observed from 2069 kg/ha to 9469 kg/ha (Fraley 2009). "Green Revolution" had also its negative impacts, such as: pollution by excessive levels of nitrogen fertilizers (7 fold increases from 1960 to 1995) and pesticides. Furthermore, monocropping had limited genetic variability contributing for vulnerability to pest and diseases threats (Tilman et al. 2002).

A new "Green Revolution" is needed when in 2050 the global population reaches 9 billion with a 50% increase from the present situation. The new "Green Revolution" needs to tackle: energy consumption, climate changes , technology, crop diversity maintenance, ecosystems biodiversity and environmental costs among others (Tilman et al. 2002; Tilman et al. 2009; Tilman et al. 2011; Ceccarelli 2012; Stamp, Visser 2012; Bellon et al. 2013; Ray et al. 2013). The target is to increase yield, but other aspects cannot be ignored such as the social component (*e.g.* traditional knowledge, smalholders role in food safety) and nutrition (Morris and Sands 2006). Indeed hunger remains related with major macronutrients (carbohydrates, fat and protein) (925 million people), but conceivably

other million suffer from 'hidden hunger', in which important micronutrients (such as vitamins and minerals) are missing with consequent risks of physical and mental impairment (Liu 2007). In contrast a billion people are substantially over-consuming, developing a new public health epidemic involving chronic conditions such as type 2 diabetes, cardiovascular disease and some cancers (Ceccarelli 2012; Tilman, Clark 2014).

Much of the responsibility for these three billion people having suboptimal diets lies within the global food system, which in turn is affected by the decreased agro biodiversity and by climate changes (Ceccarelli 2012). For this reason a new "Green Revolution" should have a holistic view, flexibility and adaptation to different circumstances (e.g adaptation to marginal areas and to polycrop systems). Additionally it integrates the participatory component for farmers, breeders and other stakeholders, promoting the participatory plant breeding and management (PPBM).

The special issue of Scientific American of August of 2013 reminded us also that food is celebration and culture, fuel and farming. The raw materials of food are genetic resources. And indeed the majority of food comes from farming using genetic resources.

I.4. Genetic Resources

From his observations of crops and their wild relative diversity, Vavilov concluded that similar patterns of variation were found

between crops and their wild relatives in unrelated crop complexes. Harlan perceived the consequences of technological and economic changes on crop diversity (Brush 2000; Maxted et al. 2002). Since 1970 substantial collection efforts were launched to hamper genetic erosion and crop vulnerability. However, conservation of crop genetic resources became independent of crop improvement.

In Portugal, Silas Pêgo understood the problem and started collection missions for maize in 1975. In the following years, a more in-depth collection supported by FAO/IBPGR in which Erna Bennet had oriented the funds (Hanelt et al. 2012) for cold storage allowed Rena Farias, as FAO consulter to cover all the country in successive missions. The collected materials, together with the previous seed stock of the Maize Breeding Station NUMI, gave rise to the first long-term cold storage facilities that were the precursors of the present Portuguese Plant Germplasm Bank (BPGV). In 2005 a collecting mission was undertaken by ITQB and IPC-ESAC (Vaz Patto et al. 2007) in the central region of Portugal and subsequent collecting missions had occurred throughout IPC-ESAC students (Santos et al. 2009) and farmers contacted under PPB research (Dinis et al. 2011).

The need of conservation led to over 1000 gene banks establishment, holding about 6 million accessions (FAO 1998). The huge amount of genetic resources poses the question of their application both to plant breeding and farming and emphasizes the gap between curators and breeders and the need of pre-breeding work. In the 1990s the understanding that on-farm conservation was in risk call

the attention of scientific community. Attitudes shift from an approach of *in situ* conservation *versus ex situ* methods to a complementary approach. It was also understood that traditional agriculture and genetic diversity were not inexorably linked and that agricultural development was not incompatible with on-farm maintenance of diversity. The awareness of scientific community for On farm conservation praxis plus participatory research led to a development of a European *in situ* (on-farm) conservation strategy from the milestone in 1992 Earth Summit in Rio and the Convention on Biological Diversity (CBD 1992), in which targeting sustainable agricultural practices that preserve natural resources, including genetic diversity, by building on enhanced agricultural research and stronger international cooperation. The European Union (EU), as CBD party, agreed that by 2020 the genetic diversity of cultivated plants, farmed and domesticated animals and of wild relatives, including other socio-economically as well as culturally valuable species, is to be maintained, and strategies have to be developed and implemented for minimizing genetic erosion and safeguarding their genetic diversity. To implement these strategies some major policy developments with impact on the conservation, use and exchange of Plant Genetic Resources for Food and Agriculture (PGRFA) (ITPGRFA, FAO 2001) and the 2nd Global Plan of Action (GPA, FAO 2011) are the most important, due to their consensus among states and cooperative nature for many European states and the European Union (EU) (Negri et al. 2015).

The resolution of the European Parliament on the EU 2020 Biodiversity Strategy also indicates that the key to the EU 2020 Biodiversity Strategy is the reform of the Common Agricultural Policy (CAP) which is "designed to support farming that ensures food safety (in a context of climate change) and promote sustainable and balanced development across all Europe's rural areas, including those where production conditions are difficult". The June 2013 reform of the CAP focused on three priorities: i) viable food production, ii) sustainable management of natural resources, and iii) balanced development of rural areas throughout the EU. Measures or programs in favor of agro-biodiversity conservation that need still to be adequately addressed by the Commission. Specifically regarding LR conservation, the Commission Directives 2008/62/EC 20 June 2008, 2009/145/EC 26 November 2009, 2010/60/EU 30 August 2010 and Commission Implementing Decision 2014/150/EU March 2014 pursuant to Council Directive 66/402/EEC on seed production and marketing opened a new way for their conservation integrated on seed production and marketing *versus* conservation *per se* (Negri et al. 2015).

When this thesis started the On-farm Conservation and Management group started as a task force. This task force becomes the On-farm Conservation and Management Working Group of the ECPGR. This group has contributed substantially for GPA implementation (Maxted et al. 2011; ECPGR 2015) and at national level a contribute to Portuguese landrace inventory (Mendes-Moreira, Veloso 2009).

The problem of germplasm evaluation by the farmers was highlighted during the "Best Ear of Sousa Valley Competition". For this competition an ear value formula was proposed by Pêgo using bibliographic data, based on high correlations with yield. The ear value formula had two applications: 1) measure the ears that farmers delivered for the competition and 2) helping farmers to select the best traits for yield improvement. Before this thesis no correlations data existed between the ears delivered for competition and their respective yield on the field. These correlations were important to better found the most adequate formula both for the competition and to improve yield. With this purpose we obtain data from the ears and respective yield, but another question occurred. What was the rank of the best ears and the rank of the best yields? Do both ranks match? What are the ear formulas that better converge both ranks?

I.5. The participatory plant breeding

PPB has been grown in the world slowly but steady grounded on scientific basis. Time is need to change mentalities and attitudes in all the participants of plant breeding, but, legislation, certification obsolete trials and institutional barriers can be other constraints. Ceccarelli (2013) refers that 47 countries had or have participatory plant breeding programs in 26 crops (13 cereals and, 6 legumes, 3 in roots or tubers, 2 in horticulture and 2 in industrial crops). It has been observed also that PPB has improved plant breeding efficiency. PPB has been considered important to tackle with climate changes

problems (Ceccarelli et al. 2013), organic and low input agriculture (Dawson et al. 2008; Dawson et al. 2013; Serpolay-Besson et al. 2014) and to polycrop systems. In addition PPB can contribute to elevate local knowledge to the role of science, community building, farmers empowerment, so as food sovereignty (Ceccarelli, Grando 2007; Machado et al. 2011). Furthermore PPB can encourage interaction between professional plant breeders, other researchers and farmers, with the objective of developing local cropping systems that better meet local needs (Cleveland 1999).

Participatory plant breeding can be also very important in on farm conservation as a source of diversity to maintain favoring dynamic gene flow between germplasm conservation and breeding (Altieri and Merrick 1987; Brush 2000; Sthapit, Friis-Hansen 2000; Cooper et al. 2001). Several approaches for increasing the diversity available to farmers have been used such as participatory varietal selection, participatory plant breeding, collaborative plant breeding and decentralized plant breeding (Cleveland 1999; Machado, Fernandes 2001; Sperling et al. 2001; Witcombe, Virk 2001; Chable et al. 2014).

To contextualize the importance of on-farm conservation and participatory plant breeding (PPB) in the late 1980s, participation has become an integrated element of sustainable development strategy and widely accepted within the United Nations and among international donor organizations. The participatory approach started to change the farming systems research on agricultural research stations throughout the inclusion of the user perspective analyses.

Participatory agricultural methodologies were refined within the CGIAR system; *e.g.*, on-farm research methodology by CIMMYT (Sthapit and Friis-Hansen 2000). Under this context, in 1984, Silas Pêgo started the VASO Project, a participatory Plant Breeding Project at Sousa Valley Region and pioneer in Europe. VASO project aims were: "How to solve the problem of the small Portuguese farmers, 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". This scientific problem had implicitly the improvement of genetic resources on-farm and on-station, preventing genetic erosion and development of methodologies for population screening and improvement. The VASO program was based on: (i) an integrant philosophy, and increasing yield without losing the parameters defined as important by the farmer, such as bread-making quality, potential for polycropping systems and use in sustainable agriculture; and (ii) the concepts of quantitative genetics in population improvement. Mass selection was applied both to landrace populations (*e.g.* 'Pigarro') and to a synthetic maize population ('Fandango'). S2 lines recurrent selection was also used in the case of 'Pigarro'. To initiate the VASO program, three main decisions had to be taken: (i) select the location that better represented both the traditional maize area and farmers interests. Indeed on this area previous agro/sociologic/economics data existed as well as the commitment and interest of local elite farmers' association (CGAVS). This context allowed the possibility to test the efficiency of an alternative project supposed to improve the

local germplasm *versus* hybrids production, at least in certain specific circumstances; (ii) select the farmers to work with – side by side, to whom the decision power would be allowed, and whose initial acceptance and enthusiasm were crucial; and (iii) select the germplasm source to start from: 'Pigarro' and 'Fandango' (Pêgo, Antunes 1997; Mendes Moreira et al. 2009). These tacit choices implied a careful respect for the local traditional agriculture. While the breeder would apply his breeding methodologies, the farmers would continue a parallel program with their own mass selection criteria. With this agreement, the breeder had to accept low input and intercropping characteristics, as well as accept and respect the local farmer as the decision maker.

Based on the concept and first year results of VASO project, Dr. Wayne Haag, as member of CIMMYT, made the decision to completely finance the VASO program that is still running.

When we started our work participatory plant breeding was mainly used in developing countries being a curiosity for developed countries (*e.g.* VASO Project in Portugal since 1984 (maize) and PPB in France since 2001 (cabbage and broccoli) or 2003 (wheat) (Chable et al. 2014). The farm seed opportunities FP6 and work done via ECPGR On-farm Conservation and Management Working Group (Veteläinen et al. 2009), COST Action 860 SUSVAR and the first Eucarpia meeting on Organic and Low input agriculture section, "Plant Breeding for organic and low-input agriculture: dealing with genotype-environment interactions in 2007, and more recently the

FP7 project SOLIBAM (2010-14), PGR-Secure (2010-14) and DIVERSIFOOD (2015-2019) will continue to expand our knowledge regarding biodiversity on farmers' fields and participatory plant breeding and management. When our work started VASO Project was running since 1984 and some evaluations were already done for 'Pigarro' maize population (Pêgo, Antunes 1997; Mendes Moreira et al. 2009). However, more recent data and with more consistency were needed, *i.e.*, trials with more environments (locations and years). This led us to enlarge the network of locations in comparison to initial VASO Project. In addition no molecular data existed that could provide us information about either diversity maintenance along farmers and breeder selection or possible important candidate genes related with particular phenotypes such as ear fasciation.

I.6. Fasciation

Fasciation describes the enlargement of the plant apex by unregulated proliferative growth (Rédei 2008; Busch, Benfey 2010) and its early description is referred by Emerson (1912). In addition fasciation is frequent in plant species (White 1948). Apparently fasciation does not confer aesthetic phenotypes. However fasciation is related with genes that alter the plant architecture and can be involved with yield. Some of these genes are *fasciated ear2* (which control the maize kernel row number) (Taguchi-Shiobara et al. 2001; Bommert et al. 2013), *fasciated ear*3 (that regulates stem cell proliferation in maize, distinct from the known CLAVATA pathway) (Je

et al. 2013), *compact plant2* (semi-dwarf plant with 1/3 height and fasciated ear) (Eckardt 2007), *double cob1* (Brewbaker 2009) and ramosa genes (*ra1*, *ra2*, and *ra3*) (Neuffer et al. 1997) that control the inflorescence branching in maize among many others (Brown et al. 2011). Other genes were recently identified, such and *fasciated ear4* (that regulates shoot meristem size in maize) (Pautler et al. 2015).

Fasciation importance was understood by some Portuguese farmers since Columbus (1492) (Ferrão 1992) till present. Indeed during the last Portuguese maize collecting expedition in 2005 (Vaz Patto et al. 2007), 56% of the traditional enduring maize landraces collected had some degree of fasciation *versus* the 10% observed during the 1980's previous collecting missions. This fact indicates farmers' preferences that can be related with adaptation to their traditional agricultural systems, *i.e.*, a germplasm with yield plasticity for different cropping systems. In fact, it has been observed that the level of expression of the fasciation trait varies with the environment, *i.e.*, more resources (*e.g.* lower densities, more nutrients and space) induce higher fasciation (Chapter IV). In addition the popular name in English, "bearsfoot" (Kempton 1923), correspond to several traditional names in Portuguese ("pe´-de-porco", "pata de porco","unha-de-porco", "mão de morto", "milho espalmado", "mãozeira" or "milho das mãozinhas") that indicates the importance of this trait for Portuguese farmers.

Embedded on Portuguese germplasm and on NUMI (Portuguese Maize Breeding Station) program, Silas Pêgo perceived the potential

of fasciation, starting its PhD studies (Pêgo 1982; Pêgo, Hallauer 1984). Pêgo's studies among other materials used the Portuguese commercial hybrid HB19 (based on the recovered WF9R and 38 11/2). Using these materials Pêgo's studies indicated that fasciation expression would be a useful trait for improving yield under specific situations of intermediate expression. For this reason fasciation expression should be considered in long term breeding programs, which would permit the proper combination of genetic factors for ear diameter, kernel row number, and ear length.

When our work started some studies on fasciation for Portuguese germplasm where already available since 1982 (Pêgo 1982) and since 2004 molecular data existed for Portuguese inbreed lines from NUMI (Vaz Patto et al. 2004), however there was neither knowledge about the percentage of fasciated maize populations that were kept by farmers nor molecular markers ever been used for Portuguese populations and fasciated material.

I.7. Aims and outline of the thesis

The ultimate goal of this work is to better understand the Portuguese germplasm, share knowledge with farmers in order that they continue to maintain and use our genetic resources being proud of them and not being poor with them. This work fits in the strategy "From kernel to bread".

In addition to an introductory **Chapter I**, the thesis comprises 6 additional chapters where the obtained results are described and discussed, plus a general discussion:

Chapter II. Participatory maize breeding in Portugal. A case study

The aim of this chapter was to present VASO project philosophy and his author motivations. In addition we wanted to know what were the main achievements' obtained in VASO Project (*e.g.* germplasm, methodologies).

Chapter III. 'Fandango': long term adaptation of exotic germplasm to a Portuguese on-farm-conservation and breeding project

With this chapter we described how 'Fandango' was created, i.e., from the 'NUTICA' development and from 'NUTICA' to 'Fandango. The other objective of this work was to survey the selection across cycles by the breeder and farmer. With this purpose, trials were conducted in Portugal and in the USA, monitoring morphological, fasciation expression, and yield.

Chapter IV. Comparison of selection methods on 'Pigarro', a Portuguese improved maize population with fasciation expression

With Chapter IV we compared the ´Pigarro´ maize OPV farmers' phenotypic recurrent selection with the breeders' S2 lines recurrent

selection response under VASO PPB Project. With this purpose, characterization at phenotypic level plus molecular data diversity of farmer selection obtained by Vaz Patto et al. (2008) was used.

Additionally, based on the evaluation trials conducted in Portugal and in USA we wanted to know which of the two selection methods was the most useful for supporting PPB in sustainable farming systems.

Chapter V. The farmers' / breeders' selection dilemma revisited by long term participatory 'Pigarro' maize breeding analysis

Regarding the comparison of ´Pigarro´ maize OPV agronomic selection response, Chapter V is an upgrade of Chapter IV, adding the evolution of molecular diversity for breeder selection plus 111000 data points to the initial 48000 measured at plot and ear level.

Besides the characterization at phenotypic and molecular level, during this long term Participatory Plant Breeding, we aim to know if the two selection methods led to the same breeding outputs; if any of the two selection methods significantly changed genetic diversity; or which of the two selection methods is the most useful for supporting PPB in sustainable farming systems.

Chapter VI. Is ear value an effective indicator for maize yield evaluation?

With Chapter VI we aim to provide tools that could help farmers on their selection, *i.e.*, converge empirical and scientific knowledge, with this purpose we target the following objectives: (1) how to develop new ear value formulas that better estimates the yield potential using ear traits. This allowed to test alternative interpretable regression methods (namely from multiple linear regression and multiple adaptive regression splines); (2) how to select the best new ear value formula to be used on ear competitions, which allowed the development of a new instance ranking method; (3) what is the adequate set of traits to select by farmers' under PPB toward better yield; and (4) what is the best Ear Value formula to use for "Sousa Valley Best Ear" competition?

Chapter VII. Genetic Architecture of Ear Fasciation in Maize (*Zea mays*) under QTL Scrutiny

Our objective with Chapter VII was to contribute to the elucidation of the genetic basis of the ear fasciation trait. Fasciation is particularly important because it is a quantitative trait that is being continuously selected by Portuguese farmers, and despite its morphological variation the impact on yield can be effective (Pêgo 1982). In this chapter we aimed to: (1) determine the genetic relationships between a comprehensive set of ear architecture traits related with fasciation in a segregating F2 population, developed from a cross between contrasting (non-fasciated PB260 x fasciated PB266) inbred lines selected in Portugal, (2) identify chromosomal positions, size

and effects of QTLs involved in the inheritance of those traits, across two environments, using univariate and multivariate approaches and (3) identify possible candidate genes associated with these QTL.

Chapter VIII. General Discussion

Finally, in Chapter VIII our aim was to integrate both phenotypic and molecular data evaluation along participatory maize breeding evolution under the VASO project. This integration was complemented with the development of a formula that could be useful for farmers' selection in a PPB methodology towards yield increase, and with the genetic basis elucidation of the ear trait fasciation, a very important ear trait to PPB farmers as a way to maintain the population resilience and yield enhancement.

I.8. References

Altieri MA, Merrick LC (1987) *In situ* conservation of crop genetic resources through maintenance of traditional farming systems. Econ Bot 41: 86–96. doi: 10.1007/BF02859354

Bellon MR, Van Etten J, Jackson M, Ford-Lloyd B, Parry M (2013) Climate change and on-farm conservation of crop landraces in centres of diversity. Plant genetic resources and climate change pp. 137-150

Bommert P, Nagasawa NS, Jackson D (2013) Quantitative variation in maize kernel row number is controlled by the FASCIATED EAR2 locus. Nat Genet 45:334–337. doi: 10.1038/ng.2534 PMID: 23377180

Brewbaker JL (2009) Double-cob (dbcb) on chromosome 1. Maize Genetics Cooperation Newsletter 1

Brown PJ, Upadyayula N, Mahone GS, Tian F, Bradbury PJ, Myles S, Holland JB, Flint-Garcia S, McMullen MD, Buckler ES (2011) Distinct genetic

architectures for male and female inflorescence traits of maize. PLoS Genetics doi: 10.1371/journal.pgen.1002383.

Brush, S. (ed.), 2000. Genes in the Field: On-farm Conservation of Crop Diversity. International Development Resources Centre/International Plant Genetic Resources Institute/Lewis, Boca Raton, Florida.

Busch W, Benfey PN (2010) Information processing without brains – the power of intercellular regulators in plants. Development 137: 1215-1226. doi: 10.1242/dev.034868

CBD (1992) Convention on Biological Diversity: Text and Annexes. Secretariat of the Convention on Biological Diversity, Montreal, http://www.cbd.int/convention/ (accessed Abr. 2015).

Ceccarelli S (2012) Plant breeding with farmers – a technical manual. ICARDA, PO Box 5466, Aleppo, Syria.

Ceccarelli S (2013) Deliverable 6.1 - Analysis of major PPB experiences worldwide. SOLIBAM - Strategies for Organic and Low-input integrated Breeding and Management. Grant agreement number: FP7- KBBE-245058 SOLIBAM

Ceccarelli S, Galie A, Grando S (2013) Participatory Breeding for Climate Change-Related Traits. In: Chittaranjan K (ed) Genomics and Breeding for Climate-Resilient Crops. Springer, pp 331-376. doi:10.1007/978-3-642-37045-8_8

Ceccarelli S, Grando S (2007) Decentralized-participatory plant breeding: an example of demand driven research. Euphytica 155: 349–360. doi: 10.1007/s10681-006-9336-8

Chable V, Dawson J, Bocci R, Goldringer I (2014) Seeds for Organic Agriculture: Development of Participatory Plant Breeding and Farmers' Networks in France. In Organic Farming, Prototype for Sustainable Agriculture. Springer Netherlands. pp. 383-400

Cleveland DA, Soleri D, Smith SE (1999) Farmer plant breeding from a biological perspective: implications for collaborative plant breeding. CIMMYT Economics Working Paper No.10. CIMMYT, Mexico, DF Cooper D, Hodgkin T, Spillane C (2001) Broadening the genetic base of crops: an overview. In: Cooper D, Hodgkin T, Spillane C (eds) Broadening the genetic base of crop production. FAO, IPGRI, CABI

Cooper D, Hodgkin T, Spillane C (2001) Broadning the Genetic Base of Crops: an Overview. In: Cooper, D., Hodgkin, T., Spillane, C. (Eds.), Broadening the genetic base of crop production. FAO, IPGRI, CABI.

Dawson J, Murphy K, Jones S (2008) Decentralized selection and participatory approaches in plant breeding for low-input systems. Euphytica 160: 143-154. doi: 10.1007/s10681-007-9533-0

Dawson JC, Serpolay E, Giuliano S, Schermann N, Galic N, Berthellot JF, Chesneau V, Ferté H, Mercier F, Osman A, Pino S, Goldringer I (2013) Phenotypic diversity and evolution of farmer varieties of bread wheat on organic farms in Europe. Genetic Resources and Crop Evolution 60: 145-163. doi: 10.1007/s10722-012-9822-x

Dias FA (2009) A importância económica das sementes e propágulos. In: Actas do I Colóquio Nacional de Sementes e Viveiros pp55-63. 29-30 Outubro. Escola Superior Agrária de Coimbra. Portugal.

Dinis I, Brites C, Santos D, Mendes-Moreira P (2011) On-farm seed production practices of organic and low-input farmers in Portugal. 20th meeting of the EUCARPIA Section Genetic Resources, Wageningen, The Netherlands.

Eckardt NA (2007) Evolution of compound leaf development in legumes: Evidence for overlapping roles of KNOX1 and FLO/LFY genes. The Plant Cell Online 19: 3315-3316

ECPGR (2015) On-farm Conservation and Management Working Group. Bioversity International - Italy. http://www.ecpgr.cgiar.org/working-groups/on-farm-conservation/ (18-11-2014).

Emerson RA (1912) Inheritance of certain “abnormalities” in maize. Am Breed Assoc Rept 8: 385–399

European Commission Report to the European Parliament, the Council and the European Economic and Social Committee - Agricultural Genetic Resources - From Conservation To Sustainable Use 2013 -COM 2013 - 838 final {SWD(2013) 486 final}. http://ec.europa.eu/agriculture/genetic-resources/pdf/swd-2013-486_en.pdf) (accessed April. 2015).

European Parliament Resolution 2012 - European Parliament Resolution of 20 April 2012 on our life insurance, our natural capital: an EU biodiversity strategy to 2020 (2011/2307(INI)). http://www.europarl.europa.eu/sides/getDoc.do?type=TA&language=EN&reference=P7-TA-2012-147)

FAO (1998) The state of the world’s plant genetic resources for food and agriculture. FAO, Rome.

FAO (2001) International Treaty on Plant Genetic Resources for Food and Agriculture. Available at: http://www.planttreaty.org/ (accessed Mar. 2015).

FAO (2011) Second Global Plan of Action for Plant Genetic Resources for Food and Agriculture, Commission on Genetic Resources for Food And Agriculture Food and Agriculture Organization of the United Nations. FAO.

Ferrão JEM (1992) A aventura das plantas e os descobrimentos portugueses. Programa Nacional de Edições Comemorativas dos Descobrimentos Portugueses, Portugal

Fraley RT (2009) Molecular genetic approaches to maize improvement–an introduction. In Molecular Genetic Approaches to Maize Improvement. Springer Berlin Heidelberg. pp. 3-6

Hanelt P, Knüpffer H, Hammer K (2012) Erna Bennett (5 August 1925–3 January 2012). Genetic Resources and Crop Evolution 59: 967-970. doi:10.1007/s10722-012-9872-0

IGC (2014) Grain Market Report. In: Report, G.M. (Ed.).

INE (2013) Previsões Agrícolas - 31 de outubro de 2013.

INE (2014) Contas Económicas da Agricultura.

INE (2014) Estatísticas do Comércio Internacional (dados provisórios, Novembro de 2014)

Je B, Lee Y, Bommert P, Komatsu M, Sakai H, Jackson D (2013) Maize Genetics Conference Abstracts., p. 214.

Kempton JH (1923) Heritable characters of maize XIV—branched ears. J Hered 14: 243-251

Liu RH (2007) Whole grain phytochemicals and health. Journal of Cereal Science 46: 207-219. doi:10.1016/j.jcs.2007.06.010

Machado AT, Fernandes MS (2001) Participatory maize breeding for low nitrogen tolerance. Euphytica 122: 567– 573. doi: 10.1023/A:1017543426136

Machado AT, Nass LL, Machado TTC (2011) Manejo sustentavel agrobiodiversidade nos biomas Cerrado e Caatinga. EMBRAPA.

Maxted N, Akparov ZI, Aronsson M, Asdal Å, Avagyan A, Bartha B, Benediková D, Berishvili T, Bocci R, Cop J, Curtis T, Daugsta K, Dias S, Duarte MC, Dzmitryeva S, Engels J, Ferant N, Freudenthaler P, Frese L, Hadas R, Holly L, Ibraliu A, Iriondo JM, Ivanovska S, Kik C, Korpelainen H, Jinjikhadze

T, Kamari G, Kell SP, Kristiansen K, Kyratzis A, Labokas J, Maggioni L, Magos Brehm J, Maloupa E, Martinez JJR, Mendes-Moreira PMR, Musayev M, Orphanidou P, Radun M, Ralli P, Sandru D, Sarikyan K, Schierscher-Viret B, Stehno Z, Stoilova T, Strajeru S, Smekalova T, Tan A, Vorosvary GM, Veteläinen M, Negri V (2011) Current and future novel threats and opportunities facing european crop wild relative and landrace diversity. In: Maxted, N., Dulloo, M.E., Ford-Lloyd, B.V., Frese, L., Iriondo, J.M. (Eds.), Agrobiodiversity Conservation. Securing the Diversity of Crop Wild Relatives and Landraces. CABI, p. 392. ISBN 9781845938512

Maxted N, Guarino L, Myer L, Chiwona E. 2002. Towards a methodology for on-farm conservation of plant genetic resources. Genet Resour Crop Ev 49: 31-46. doi: 10.1023/A: 1013896401710

Mendes Moreira P, Veloso MM (2009) Landrace Inventory for Portugal. In: Veteläinen, M., Negri, V., Maxted, N. (Eds.), European landraces onfarm conservation, management and use. Bioversity International, Rome, Italy, pp. 124-136.

Mendes Moreira PMR, Pêgo S, Vaz Patto MC (2009) On-farm Conservation Portugal – Vaso Project – a Long-term Conservation Programme. In: Veteläinen M, Negri V, Maxted N (Eds.), European landraces onfarm conservation, management and use. Bioversity International, Rome, Italy, pp. 275-282.

Morris CE, Sands DC (2006) The breeder's dilemma yield or nutrition? Nat Biotech 24, 1078-1080.

Negri V, Freudenthaler P, Gasi F, Goldringer I, Heinonen M, Maxted N, Mendes Moreira P, Strãjeru S, Tan A, Torricelli R, Veteläinen M, Voegel R, Weibull J (2015) A European *In situ* (On-Farm) Conservation Strategy for Landraces. CAB (in press)

Neuffer MG, Coe EH, Wessler SR (1997) Mutants of maize, Cold Spring Harbor Laboratory Press, NY.

Pautler M, Eveland AL, LaRue T, Yang F, Weeks R, Lunde C, Je BI, Meeley R, Komatsu M, Vollbrecht E, Sakai H, Jackson D (2015) FASCIATED EAR4 Encodes a bZIP Transcription Factor That Regulates Shoot Meristem Size in Maize. The Plant Cell Online. doi: 10.1105/tpc.114.132506

Pêgo SE (1982) Genetic potential of Portuguese maize with abnormal ear shape, Ph.D. Thesis, Iowa State Univ.

Pêgo SE, Antunes MP (1997) Resistance or tolerance? Philosophy, may be the answer. In: Proceedings of the XIX – Conference of the International

Working Group on Ostrinia. Guimarães Portugal 30th August–5th September 1997

Pêgo SE, Hallauer AR (1984) Portuguese maize germplasm with abnormal ear shape. Maydica 29: 39–53

PGR Secure (2015) Introduction to PGR Secure. http://www.pgrsecure.org/ (accessed April. 2015).

Ray DK, Mueller ND, West PC, Foley JA (2013) Yield Trends Are Insufficient to Double Global Crop Production by 2050. PLOS ONE 8, e66428. doi: 10.1371/journal.pone.0066428

Rédei GP (2008) Fasciation. Encyclopedia of Genetics, Genomics, Proteomics and Informatics. Springer Netherlands, pp. 673-673.

Santos JPN, ESAC-Students, Brito O, Tristany M, Veloso M, Santos D, Vaz Patto MC, Mendes-Moreira P (2009) Genetic Resources Conservation and Green Care in Agriculture: a link among institutions. Farming for Health, Pisa, Italy.

Serpolay-Besson E, Giuliano S, Schermann N, Chable V (2014) Evaluation of Evolution and Diversity of Maize Open-Pollinated Varieties Cultivated under Contrasted Environmental and Farmers' Selection Pressures: A Phenotypical Approach. Open Journal of Genetics 4: 125-145. doi: 10.4236/ojgen.2014.42014

SOLIBAM (2015) Strategies for Organic and Low-input Integrated Breeding and Management. Collaborative Project. http://www.solibam.eu/modules/addresses/viewcat.php?cid=1. (accessed April. 2015).

Sperling L, Ashby JA, Smith ME, Weltzien E, McGuire S (2001) A framework for analyzing participatory plant breeding approaches and results. Euphytica 122: 439-450. doi: 10.1023/A:1017505323730

Stamp P, Visser R (2012) The twenty-first century, the century of plant breeding. Euphytica 186: 585-591. doi: 10.1007/s10681-012-0743-8

Sthapit B, Friis-Hansen E (2000) Concepts and rationale of participatory approaches to conservation and use of plant genetic resources. In: Friis-Hansen E, Sthapit B (eds) Participatory approaches to the conservation and use of plant genetic resources. International Plant Genetic Resources Institute, Rome Italy

Taguchi-Shiobara F, Yuan Z, Hake S, Jackson D (2001) The fasciated ear*2* gene encodes a leucine-rich repeat receptor-like protein that regulates

shoot meristem proliferation in maize. Genes Dev 15: 2755-2821. PMID: 11641280

Tilman D, Balzer C, Hill J, Befort BL (2011) Global food demand and the sustainable intensification of agriculture. Proceedings of the National Academy of Sciences 108: 20260-20264. doi: 10.1073/pnas.1116437108

Tilman D, Cassman KG, Matson PA, Naylor R, Polasky S (2002) Agricultural sustainability and intensive production practices. Nature 418: 671-677. doi: 10.1038/nature01014

Tilman D, Clark M (2014) Global diets link environmental sustainability and human health. Nature 515: 518–522. doi:10.1038/nature13959

Tilman D, Socolow R, Foley JA, Hill J, Larson E, Lynd L, Pacala S, Reilly J, Searchinger T, Somerville C, Williams R (2009) Beneficial Biofuels—The Food, Energy, and Environment Trilemma. Science 325: 270-271. doi: 10.1126/science.1177970

Trewavas A (2002) Malthus foiled again and again. Nature 418: 668-670. doi: 10.1038/nature01013

Vaz Patto MC, Moreira PM, Almeida N, Šatović Z, Pêgo S. 2008. Genetic diversity evolution through participatory maize breeding in Portugal. Euphytica 161:283-291. doi: 10.1007/s10681-007-9481-8

Vaz Patto MC, Moreira PM, Carvalho V, Pêgo S (2007) Collecting maize (*Zea mays* L. convar. *mays*) with potential technological ability for bread making in Portugal. Genet Res Crop Evol 54:1555‑1563. doi: 10.1007/s10722-006-9168-3

Vaz Patto MC, Šatović Z, Pêgo S, Fevereiro P (2004) Assessing the genetic diversity of Portuguese maize germplasm using microsatellite markers. Euphytica 137: 63-72. doi:10.1023/B:EUPH.0000040503.48448.97

Veteläinen M, Negri V, Maxted N (2009) European landraces onfarm conservation, management and use. Bioversity International, Rome, Italy

White OE (1948) Fasciation. Bot Rev 14: 319-358

Witcombe JR, Virk DS (2001) Number of crosses and population size for participatory and classical plant breeding. Euphytica 122: 451-462. doi: 10.1023/A:1017524122821

CHAPTER II.

Participatory maize breeding in Portugal. A case study

The work presented in this chapter was mostly performed by Pedro Mendes Moreira (see acknowledgments section) and partially included in the following publication: MENDES MOREIRA, P. M. R. (2006). Participatory maize breeding in Portugal. A Case Study. Acta Agronomica Hungarica, 54(4): 431-439.

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
/UNL

Knowledge Creation

II.1. Abstract

Participatory maize breeding (PMB) was initiated in Portugal in 1984 by Dr. Silas Pêgo at Sousa Valley. The VASO project was intended to answer the problem facing small farmers, *i.e.* yield increasing without losing the parameters defined by farmers in polycropping systems maintaining the quality traits under a sustainable agriculture. This model is based on the Integrant Philosophy, which contrasts with the Productivist Philosophy. The Integrant Philosophy is intended to fit a multicrop agricultural system that corporate agriculture does not reach due to incipient market conditions. The present document intends to be a contribution to: 1) the study of 20 years of VASO; 2) methods used in PMB for Portuguese open-pollinated maize varieties and 3) present research.

II.2. Introduction

Twenty years have passed since the beginning of the Sousa Valley Project (VASO) in 1984. This paper is intended as a contribution to the evolution of maize breeding and genetic resources in Portugal, and intends to stress the importance of 20 years of participatory maize breeding (PMB) in the Portuguese Northern Sousa Valley region. Any description of VASO must be closely connected with Dr. Silas Pêgo, the founder of the Integrant Philosophy approach, which had its practical application through on-farm breeding. Pêgo also conducted the first basic implementation of the Portuguese Plant Gene Bank (BPGV). For a better understanding of these achievements

some biographic data will be presented. An overview of this project will also be provided.

Silas Pêgo is the kind of scientist who always thinks of science as a means to directly benefit farmers. His career, as well as his life, was early connected with maize. Born to Bento Fernandes Pêgo and Maria Esteves Pêgo in June 1942 in a small farming community in the extreme North of Portugal (Pias, Monção), he likes to say that he was born 50 m away from a maize field. A farmer's son, he grew up on a small farm in Minho province where polycrop systems are usual. These facts were crucial in his rethinking of the relationship between breeder and farmer. He graduated at Instituto Superior de Agronomia, Universidade Técnica de Lisboa in 1972. He started his professional career at Estação Agrária de Braga in its Núcleo de Melhoramento de Milho (NUMI) (maize breeding centre in Braga city). During his work at Braga he took several courses at DG/EAN (Genetics Department of National Agronomic Station, Oeiras) under the guidance of Professor Miguel Mota, who was responsible for the theory behind the Nutica population, a germplasm basis that Pêgo would use as a precursor of 'Fandango' (one of the biggest ear-size germplasms in the world).

Later on, as director of NUMI, he laid the foundations of the future Portuguese Plant Gene Bank (BPGV), which was responsible for the Mediterranean Programme of FAO/IPGRI. He also organized and participated in several national and international germplasm collecting missions.

At NUMI he continued the research of António Lacerda, predecessor of Luís Freire de Andrade, his peer. He observed that some pure lines with fasciation expression showed several problems of stabilization. This problem was then used for his PhD research thesis in Plant Breeding and Cytogenetics at Iowa State University (ISU), USA, concluded in 1982 (Pêgo 1982). The research developed by Pêgo under the supervision of Prof. Arnel Hallauer was a unique work done with Portuguese germplasm in the USA, and is still a hallmark for those who intend to work on maize fasciation (Pêgo, Hallauer 1984). Before presenting his thesis he received a congratulatory mention from his advisor for his discovery of the U gene. As a scholar of The Rockfeller Foundation, before leaving the USA he obtained the permission of the foundation to extend the scholarship in order to discuss a maize breeding programme for Portuguese conditions with his former professors.

How to solve the problems facing small Portuguese farmers, where land is scarce and population density is high, *i.e.* where the American agriculture model is not appropriate and where the multinationals do not have a market to operate in, was another issue that encouraged him to conduct further research. From 1982 to 1985, Silas Pêgo was responsible for the Maize National Programme and, together with his mentor, Dr. Luís Costa Rodrigues, organized and constructed the National Breeding Programme, with two main components: 1) On-station approach, 2) On-farm approach, *i.e.* a Monoculture System (hybrid programme), adapted to the Productivist Philosophy, and

Polycrop Systems (breeding populations), adapted to the Integrant Philosophy (Pêgo, Antunes 1997).

The PMB Programme, as an Integrant Philosophy approach, was initiated in one of the best locations, side-by-side with Lousada farmers. The multidisciplinary scientific team attracted CIMMYT support from 1985 until Portugal joined the European Community.

Integrant Philosophy and Productivist Philosophy are not necessarily antagonists. Integrant Philosophy could be a very effective method of achieving diversity and germplasm for the Productivist Philosophy. According to the research done by Hallauer during the 70s and 80s, his populations began to be more productive than otherwise comparable commercial hybrids. The inbred lines obtained from these populations led to a new generation of better performing hybrids, *i.e.* from new improved populations it has been possible to extract superior inbred lines responsible for a continued rise in maize yield. Several authors (Altieri and Merrick 1987; Brush 1995; Bellon 1996; Jarvis, Hodgkin 1998; Sthapit et al. 2005) have focused on the importance of *in situ* conservation as a source of diversity to maintain a dynamic gene flow between germplasm conservation and breeding. This scientific rationality not only constitutes the basis for Pêgo's suggestion that the VASO project should be repeated in several regions of the country, but also stresses the importance of the pre-breeding approach, another of Pêgo's research topics, in which he developed some straightforward methods for germplasm evaluation. As Pêgo stresses, the importance of pre-breeding is related with the

need to reduce the gap between "curators" and "breeders" or between "characterisation" and "utilisation". In fact, genebank catalogues represent a huge amount of data, as the IPGRI list of passport data parameters, but the most important ones for breeders, due to their direct relation with yield – inbreeding depression, combining ability and stress behaviour – are missing. If a breeder could afford to have even a preliminary evaluation of such parameters, this would allow him to screen a vast set of accessions for those with a better chance of success. Some examples of these proposed methodologies are discussed in Overlap Index Method (Moreira, Pêgo 2003) and “HUNTERS” (Moreira et al. 2005a).

The Integrant Philosophy model, elaborated by Pêgo in 1983, was the approach used to tackle the reality facing, small farmers in Portugal where arable land is scarce and the population density is high. Under these small plot conditions the American model does not give an appropriate answer and the multinationals do not have attractive market conditions. The Integrant Philosophy approach takes into account not only the agricultural system, but also the farmer, as the most important genetic resource with the power of decision (Table II.1). Pêgo’s Integrant Philosophy is also the result of background interaction between: agriculture on small plots of land, the importance of genetic resources in breeding, an overview of maize in the world (FAO consultant), population improvement methodologies and the NUMI hybrid program.

Table II.1Contrasting issues and/or consequences between the two philosophical models: productivist versus integrant (Pêgo, Antunes,1997)

Contrasting factors	**Philosophical model**	
	Produtivist	**Integrant**
1. Profession of faith	Yield is the determinant factor	Farmer's decisions are rational
2. Decisive centre	The seed (breeder)	The farmer
3. Dynamic action	Centripetal	Centrifuge
4. Energy	Fossil	Renewable
5. Row materials	Exotic, inbreeds	Local adapted populations
6. Science		
6.1. Gene action	Non-additive (heterosis)	Mainly additive
6.2. Breeding methods	Genealogical selection (+) biotechnology	Recurrent selection (-) biotechnology
6.3. Pathology	Resistance	Tolerance
6.4. Technology	(+) Mechanization (+) agrochemical (-) manpower and monocropping	(-) Mechanization (-) agrochemical (+) manpower and polycropping (system)
7. Type of seed	Hybrid, uniformity	Open-pollinated, diversity
8. Final output	High yielding, quantity	Moderate yielding, quality
9. Environmental effects		
9.1. Protection level	Soil, water and air pollution	Soil, water and air cleanness
9.2. Genetic resources	Erosion	Conservation
9.3. Farming continuity	Leading to exhaustion	Sustainability

II.3. Results

Together with the on-farm project conducted in Seropédica in Rio de Janeiro State, Brazil (Machado and Fernandes 2001), VASO started in 1984. Nevertheless, the VASO project in Lousada is probably the oldest PMB project in the world, because it has maintained, from the very beginning, different sets of germplasm identified and conserved under cold storage conditions. As an overall summary its output has resulted in the following improved populations:

Pigarro (FAO 300 white flint), Amiudo (FAO 200 yellow flint), Aljezur (FAO 400 yellow flint), Aljezudo (FAO 300 yellow flint), Castro verde (FAO 600 yellow flint) and 'Fandango' (FAO 600 yellow dent).

During the 2005 season, the evaluation of sets representing the distribution in time over the 20 years, cycles of phenotypic recurrent selection ('Pigarro' and Fandango) and S2 lines recurrent selection ('Pigarro') were carried out in three locations in Portugal and 5 locations in Iowa State, USA, and other evaluation sets are still underway. Nevertheless, some prior analyses have already been published (Pêgo, Antunes 1997), yielding the following information:

1 - 'Pigarro' produces tall plants with high ear placement and a high level of ear fasciation, responsible for a large number of kernel rows and consequently an improved kernel weight per plant.

2 - A gain of 17% (genetic and environmental) was registered when a comparison was made between C084 (7.0 $Mgha^{-1}$) and C1-S2 (8.2 $Mgha^{-1}$).

3 - Significant differences were detected between both C1 and C086 and C090, but no significant differences were observed between the C0s.

4 - The analysis of data on stalk and root lodging showed that the best yields depended on a combination of large ear size and good stalk and root characteristics.

5 - The evolution of phenotypic recurrent selection, from 1985 to 1990, did not lead to significant differences, but a positive tendency was registered (2% between C086–C084 and 2.4% between C090–C086).

6 - In plant quality and pest tolerance control, the farmer found somewhat contradictory results for root and stalk lodging between the first (84–86) and second (86–90) periods. This circumstance illustrates the communication and acceptance between farmer and breeder, discussed by Pêgo and Antunes (1997), and is a very interesting sociological testimony that stresses the importance of the breeder–farmer relationship and who really makes the decisions!

However, one major aspect of this project is linked with international evaluation. At the beginning of the PMB project in Sousa Valley, Dr. Wayne Haag (as CIMMYT director for maize breeding in the Mediterranean area), after having observed 'Fandango' in the field, asked, "Where do we in America have an open-pollinated population like this, yielding 10 tonnes per ha?". As an immediate consequence, he decided to link CIMMYT with this project by supporting both its logistics and finances from 1985 until Portugal entered the European Economic Community (EEC).

In 2004 Professor Arnel Hallauer visited the project and after maize field observations he also mentioned in his report, "…In addition to reviewing the program with Dr. Silas Pêgo, I also had the opportunity to visit the farm of Mr. Francisco Ribeiro Meireles… Maize growth on the farm, and surrounding areas, looked very good. It seems good to excellent yields can be expected for that particular area".

Finally, the recent introduction of 'Pigarro' in the central province of Huambo, through the initiative of the Angolan, governmental authorities, completes the picture. This improved open-pollinated

white-flint variety – the preferred type of maize for food (Hallauer 2004) – was chosen for its bread quality. Due to its good adaptation in the first year, multiplication facilities were built in Angola in order to supply small-scale African farmers to improve their living standard – one of the two aims for which VASO was born!

II.4. Discussion

From the beginning of VASO (1985) till the present time, the breeding process has been continued with the initial germplasm basis. The results presented in Pêgo and Antunes (1997), referring to breeding population methodologies that favour diversity and tolerance, indicate that non-adaptation to the competitive models of production imposed by the hybrid industry cannot be applied in all circumstances. It is strongly recommended that these two systems should work side by side because, besides giving a direct response to the problems facing small, quality-oriented, sustainable farming, the Integrant Philosophy also offers new germplasm sources for the hybrid industry, which is always eager for new inputs of improved genetic bases from which new inbreds can be extracted. In other words, the Integrant Philosophy could also be an important complement to the Productivist Philosophy, if more research on prebreeding - an area that needs an effective approach between genetic resources and breeding – is done.

The VASO project suggests that this scientific approach should be replicated in several places in the country, especially in mountainous

areas, where *in situ* conservation and sustainable quality-oriented agriculture could work together as part of a rural developmental policy, thus framing the economic basis for small-farming communities. As extra outputs, new improved sources of quality-oriented germplasm could also serve the hybrid seed industry.

It is our opinion that, for the present and future, Portugal could play an important role in on-farm conservation, especially in white-flint maize, due to its traditional diet, probably unique in the world, based on maize bread ("Broa"). Even in the 21st century, maize could still have a say in the economic recovery of Portuguese organic farming. And if a greater role can be played in Africa, let the good news be spread to wherever it is needed!

II.5. Material and Methods

VASO was implemented according to the Integrant Philosophy point of view. To achieve this goal three main decisions were taken: 1) The choice of location to represent the region, 2) the germplasm to start from, and 3) the farmer to work side-by-side with (Pêgo, Antunes 1997).

II.5.1 Location

The Sousa Valley was chosen, taking into account the following factors: (a) Location in a traditional maize area characterized by polycropping systems, where maize still plays an important role, (b)

One of the most fertile areas in the Northwest region of Portugal, (c) In 1985, 20–25% of its soils were planted with hybrids, compared with 15% as a national average. It was also on this area that the maize production (18 $Mgha^{-1}$) champion was located (Mr. Coreolano), (d) The availability of a basic amount of agro/sociologic/economic data previously collected by members of the original multidisciplinary team provided the breeder with a systemic knowledge of the region, (e) The support of a local elite farmers' association (CGAVS) which agreed to be part of the project, (f) The possibility to test the efficiency of an alternative project expected to improve the local germplasm in order to be competitive, at least under certain specific circumstances, sideby- side with the local farmers.

II.5.2 Local germplasm

One of the pre-requisites of the Integrant Philosophy option (Table II.1) was the existence of local adapted germplasm. This option respects the farmers' selection pursued over the last four centuries and also assures the environmental adaptation already achieved either for the soil/climate or for quality preferences. This assumption led to an extensive survey in the Sousa Valley Region, in the summer of 1984, looking for the best open-pollinated varieties (OPV) in the field. This survey allowed a reasonable choice of germplasm to start from: two OPVs were chosen, an early yellow flint variety (FAO 200) adapted to stress conditions (Al toxicity and water limitations) known as Amiudo, and a white flint medium maturity variety (FAO 300) with

strong fasciation expression. Both varieties showed a high percentage of stalk and root lodging, like the great majority of landraces. Prior selection was made according to: second class soils (the first quality soils were already reserved for competitive hybrids), low nitrogen inputs, water limitations, flint type kernel, bread-making quality selected by the farmers, and polycrop system integration (maize-beans-*Lolium* sp.). This regional white flint OPV was named 'Pigarro', after an agreement between farmer and breeder.

II.5.3 Exotic germplasm

Fandango (FAO 600) is an open-pollinated selected composite derived from Nutica following the Design I crossing methodology. The Nutica broad population (FAO 700) was composed by intercrossing 76 yellow (dent and flint) elite inbred lines from the NUMI programme in natural isolation. In this set of 76 inbreds, 20% were Portuguese germplasm and 80% American germplasm.

The preparation of the material to be included in Nutica began in 1974. The Nutica Project was initiated in 1975 and finished in 1978.

In 1983, after Pêgo's return from the USA, the latest version of Nutica (now almost entirely yellow dent) was included in his program at ENMP (Elvas Breeding Station). In 1984, with the purpose of evaluating the gene action composition, the population was submitted to crosspollination, type Design 1 (1 male crossed with 5 females), as part of the MSc project of Fátima Quedas under the

supervision of Pêgo. The results obtained in the 2nd year trial were very promising, with high yielding levels obtained in the borders (composed by all the crosses in the trials). Due to the isolation conditions of the field, Pêgo used a mixture obtained in open pollination as a first basis of what would be designated as Fandango.

In 1985 Pêgo introduced 'Fandango' in Lousada and stratified mass selection has been applied since then. This 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 13 years, 'Fandango' has been the winner of the contest "Best ear of Sousa Valley Region".

II.5.4 The farmer

Choosing the right people to work with is also a major decision in an on-farm project, where the work is carried out side-by-side with the farmer himself, to whom the power of decision will be delegated. All the information gathered was decisive for the choice of the two farmers. Their initial acceptance and enthusiasm to join the project turned out to be the best guarantee of success.

So, with careful respect for the local traditional agriculture, a deal was made with the farmers involved: while the breeder would apply his breeding methodologies, they should continue a parallel programme with their own mass selection criteria. With this tacit deal between breeder and farmer, three consequences became

clear: 1) Respecting the “system” would imply accepting low input and intercropping characteristics, as well as accepting and respecting the local farmer as the decision maker, 2) With two simultaneous breeding programmes (the farmer’s and the breeder’s) the farmer would have a constant possibility to compare the effectiveness of both. This would allow the farmer to base his decisions on solid grounds, and 3) The option of diversity and quality as the first priority trait, due to starting from local adapted germplasm.

II.5.5 Breeding methodologies

In order to address both yield component and pest and diseases problem, the breeding approach was to use quantitative genetics through population improvement selection, combining three main recurrent selection methodologies: phenotypic, S1 and S2 lines (Pêgo, Antunes 1997).

II.5.6 Phenotypic recurrent selection

This methodology, involving mass selection with a two-parent control, is an improved extension of the common mass selection usually performed by all farmers (with only one parent control) and is the breeding tool lately used by the farmer, who has been advised to carry it out in a three-step sequence (A–B–C), the first two steps (A and B) in the field and the third one (C) during storage. The sequence follows this pattern:

Immediately before pollen shedding, selection is performed for the male parent by detasselling all the undesirable plants (pest and disease susceptible, weakest, plants that do not fit the desirable ideotype).

Some days before the harvest, besides selecting for the best ear size, the plants are kicked at their base (first visible internodes) to evaluate both their root and stalk quality. And, as an indirect measurement, the pest and disease tolerance can also be evaluated. In practical terms, if the plant does not resist the impact and lodges, it is eliminated. Moreover, special preference in selection is given to prolific plants.

In storage, after harvest, selection is performed separately for normal and prolific ears and always includes, besides ear length and kernel row number, prolificacy, and the elimination of damaged/diseased ears. The selected ears from both sets are finally shelled and mixed together to form the next generation seed.

II.5.7 Recurrent selection of S1 and S2 lines

Selection based on S2 lines was initially the method to be applied to the two chosen regional germplasms (‘Pigarro’ and ‘Amiudo’) due to its good indication of the additive component of genetic variance (3/2σa 2) (Hallauer 1992). Nevertheless, while ‘Pigarro’ could be selfed well up to the S2 stage, ‘Amiudo’ exhibited such strong inbreeding depression that normal yield tests on S2 lines became

impossible. As a consequence, the yield tests were conducted on remnant S1 seed according to the S1 Lines Recurrent Methodology.

In the S2 lines option, 1000 S1 lines and then 500–600 S2 lines were selected. The next step was the selection of 200 S2 lines to be used in a yield trial, where 15 to 20% selection pressure was applied and a final set, *i.e.* 30–35 elite S2s, was selected for the recombination season in order to form the first cycle seed (C1), and so on. During the selection process, the selection of plants to be selfed and selection before harvest led to the systematic elimination of diseased plants.

II.6. Acknowledgements

Article drafting by PMM

Paper critically reviewed by SP

II.7. References

Altieri MA, Merrick LC (1987) *In situ* conservation of crop genetic resources through maintenance of traditional farming systems. Econ Bot 41: 86–96. doi: 10.1007/BF02859354

Bellon MR (1996) The dynamics of crop intraspecific diversity: a conceptual framework at the farmer level. Econ Bot 50:29–36

Brush SB (1995) *In situ* conservation of landraces in centres of crop diversity. Crop Sci 35:346–354. doi:10.2135/cropsci1995.0011183X003500020009x

Hallauer AR (1992) Recurrent selection in maize. In: Janick, J. (ed.), Plant Breeding Reviews. John Wiley & Sons, Inc., New York, pp. 115–177.

Hallauer AR (2004) Specialty corns. In: Smith, C. W. (ed.), Corn: Origin, History, Technology, and Production. John Wiley & Sons, Inc., New York, pp. 897–933.

Jarvis D, Hodgkin T (1998) Strengthening the scientific basis of *in situ* conservation of agricultural biodiversity on-farm. Options for data collecting and analysis. Proc. of a workshop to develop tools and procedures for *in situ* conservation on-farm. 25–29 August 1997, Rome, Italy.

Machado AT, Fernandes MS (2001) Participatory maize breeding for low nitrogen tolerance. Euphytica 122: 567– 573. doi: 10.1023/A:1017543426136

Moreira PM, Pêgo S (2003) Pre-breeding evaluation of maize germplasm. The case of a Portuguese open-pollinated variety. In: AR Hallauer (ed.), Proceedings of the International Symposium on Plant Breeding. Mexico City, Mexico, 17–22 August, 2003.

Moreira PM, Santos JP, Simões P, Santos JP, Vaz Patto MC, Carvalho V, Pêgo S (2005a) Pré-avaliação de populações de milhos regionais da região centro. A utilização do método «HUNTERS». II Colóquio de Melhoramento de Plantas e Conservação de Recursos Genéticos. Santarém 18 de Novembro. Escola Superior Agrária de Santarém.

Moreira PM, Santos JP, Simões P, Santos JP, Vaz Patto MC, Carvalho V, Pêgo S (2005b) Pré-avaliação de Populações de Milhos Regionais da Região Centro. Parâmetros Biométricos e Fitossanitários. VII Encontro Nacional de Protecção Integrada. 6 a 7 de Dezembro. Coimbra.

Pêgo SE (1982) Genetic potential of Portuguese maize with abnormal ear shape, Ph.D. Thesis, Iowa State Univ.

Pêgo SE, Antunes MP (1997) Resistance or tolerance? Philosophy, may be the answer. In: Proceedings of the XIX – Conference of the International Working Group on Ostrinia. Guimarães Portugal 30th August–5th September 1997

Pêgo SE, Hallauer AR (1984) Portuguese maize germplasm with abnormal ear shape. Maydica 29: 39–53

Sthapit B, Sajise P, Jarvis D (2005) Community based on-farm conservation of agricultural biodiversity: Good practices and lessons learned from Nepal and Vietnam. Second Colloquium on Plant Breeding and Plant Genetic Resources Conservation organised by Portuguese Association of Horticulture (APH) in Santarém City, Portugal, 18 November 2005

CHAPTER III.

‘Fandango’: long term adaptation of exotic germplasm to a Portuguese on-farm-conservation and breeding project

The work presented in this chapter was mostly performed by Pedro Mendes Moreira (see acknowledgments section) and partially included in the following publication: Mendes Moreira, P., Vaz Patto, M.C., M. Mota, M., J. Mendes-Moreira, J., Santos, J.P.N., Santos, J.P.P., Andrade, E., Hallauer, A.R., Pêgo, S.E., 2009. ‘Fandango’: long term adaptation of exotic germplasm to a Portuguese on-farm-conservation and breeding project. Maydica 54, 269-285.

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
/UNL
Knowledge Creation

III.1. Abstract

This study presents:

I - The two steps genesis of the synthetic maize population 'Fandango'. A) 'NUTICA' creation: in 1975, Miguel Mota and Silas Pêgo, 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 Pêgo, 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.

III.2. Introduction

Sustainability in agriculture emphasizes 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 16th 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 (Pêgo, 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 Pêgo 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 (Pêgo, 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 (Pêgo, 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.

III.3. Results

III.3.1 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 plant 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.*, plant 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 III.1).

Mass Selection at Iowa - The regression analysis conducted to estimate direct response to selection revealed significant decrease for yield (Table III.1, Table III.2). Greater proportion of the variation was explained 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 III.1).

MARS analysis showed no variation across cycles of selection for root and stalk lodging. (Table III.2).

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.

Table III.1 After positive assumption of normality, linear regression was used. Estimation of linear regression coefficient (b), their standard errors, initial cycle prediction (Ĉ0), coefficients of correlation (R) 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.

Traits	Mass selection Iowa													Populations Standard Iowa			
	b				Ĉ0	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
Traits	Mass selection Portugal													Populations Standard Portugal			
	b				Ĉ0	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
Rachis 1, cm	0.03	±	0.00	**	2.35	0.97	1.15	**	**	**	*	**	**	2.47	2.02	2.16	1.93
Rachis 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 plant stand correspond to the average of the correspondent cycles.

%Gain/Y – percentage of gain per year, ANOVA for C-cycles of selection, E-environment; Years; x-interactions; Ĉ - predicted cycle of selection, except for plant stand that was calculated the average. Flowering data was not measured in Lousada, Portugal in 2008. Shaded portions distinguished were *Analysis* of Variance was not done from the white portions were non significant differences were registered.

Table III.2 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	TEPRE C6
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
A**N**gle	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-	1.77	1.11	1.07	1.07

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	TEPRE C6
		19)}				
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.

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) (Figure III.1, Figure III.2). 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) (Table III.1, Table III.2, Table III.3, Table III.4).

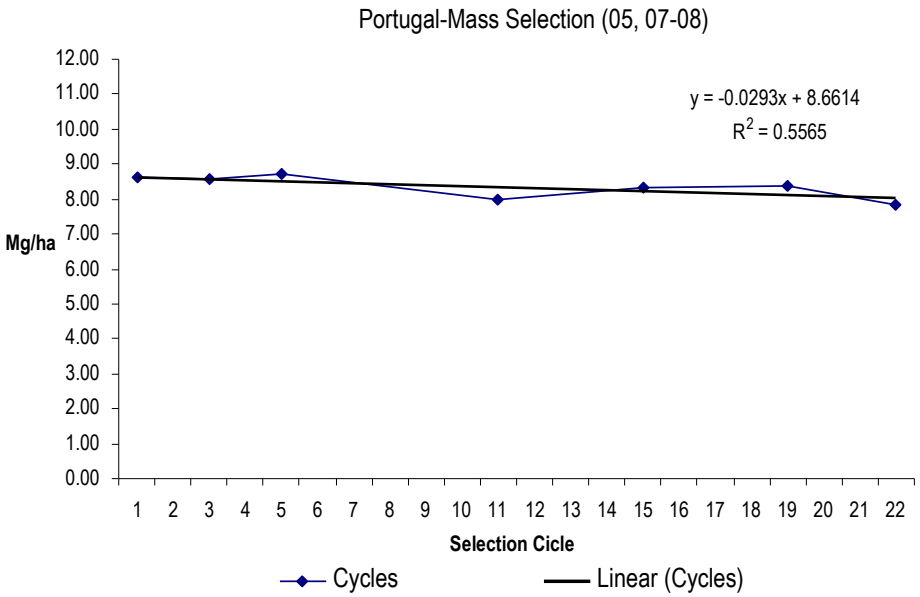

Figure III.1 Yield evolution during the 22 cycles of mass selection.

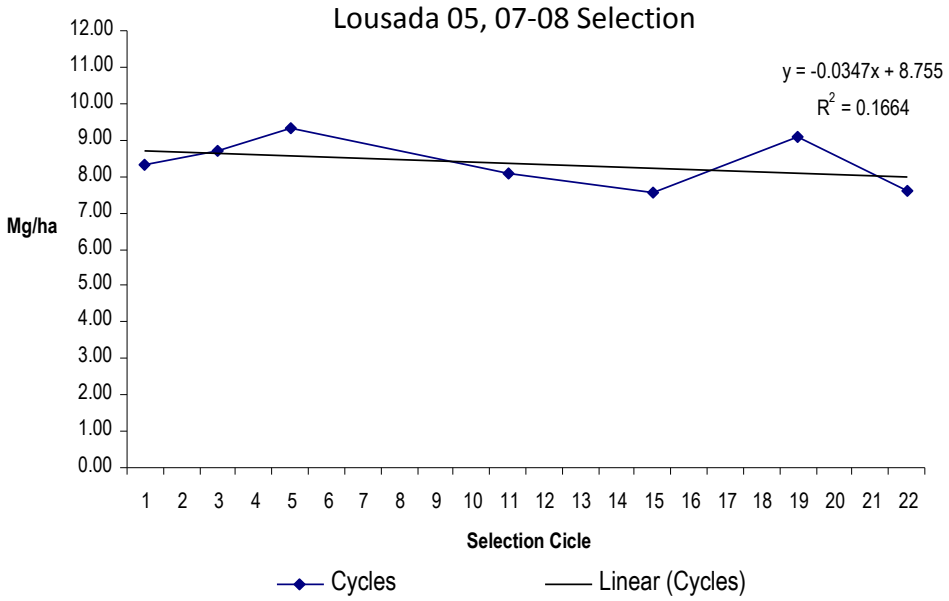

Figure III.2 Yield evolution during the 22 cycles of mass selection for Lousada. The first five cycles represent the breeder selection.

Table III.3MARS 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 original cycle (mean trait in bold) plus the transformation.

Table III.4 After positive assumption of normality, linear regression was used. Estimation of linear regression coefficient (b), their standard errors, initial cycle prediction (Ĉ0), coefficients of correlation (R) 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.

Traits	Mass selection Pt - Lousada										Populations Standard Pt -Lousada			
	b				Ĉ0	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
Rachis 1, cm	0.03	±	0.00	**	2.38	0.99	1.18	**		**	2.43	2.00	2.23	2.03
Rachis 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; Years; x-interactions; Ĉ - predicted cycle of selection, except for plant stand that was calculated the average. Flowering data was not measured in Lousada, Portugal in 2008.

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 III.3, Table III.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 III.2). 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 III.3).

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 kernelrow- 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 (Table III.1, Table III.2, Table III.3, Table III.4).

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 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 Pêgo, 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 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 (Table III.1, Table III.2, Table III.3,Table III.4).

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, plant 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 III.7 and Figure III.3, Figure III.4 and Figure III.5.

Table III.5 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-ED1)}	
	-0.0134494*max{0,(SW-328.499)}	
	-0.5679192*max{0,(6-N)}	
	+2.978107*max{0,(ED4-4.515)}	
	+3.87864*max{0,(1.35-Fa)}	
	-4.968153*max{0,(ED3-4.94)}	
	+7.738557*max{0,(ED4-5.095)}	
	+2.688185*max{0,(M1-1.4075)}	
	+0.3699823*max{0,(U-2)}	

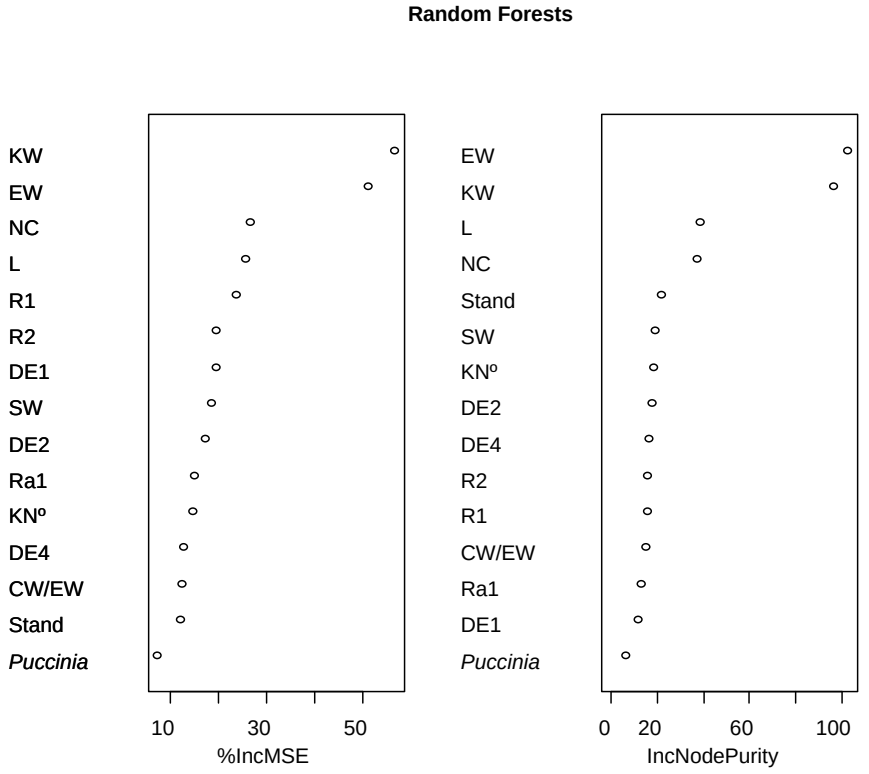

Figure III.3 Mean Decrease Accuracy (%IncMSE) and Mean Decrease MSE (IncNodePurity): there is no clear guidance on which measure to prefer (Kuhn et al., 2008). The independent variable is Yield. They are presented only the 15 most relevant dependent variables. The percentage of variation explained was 46.7%.

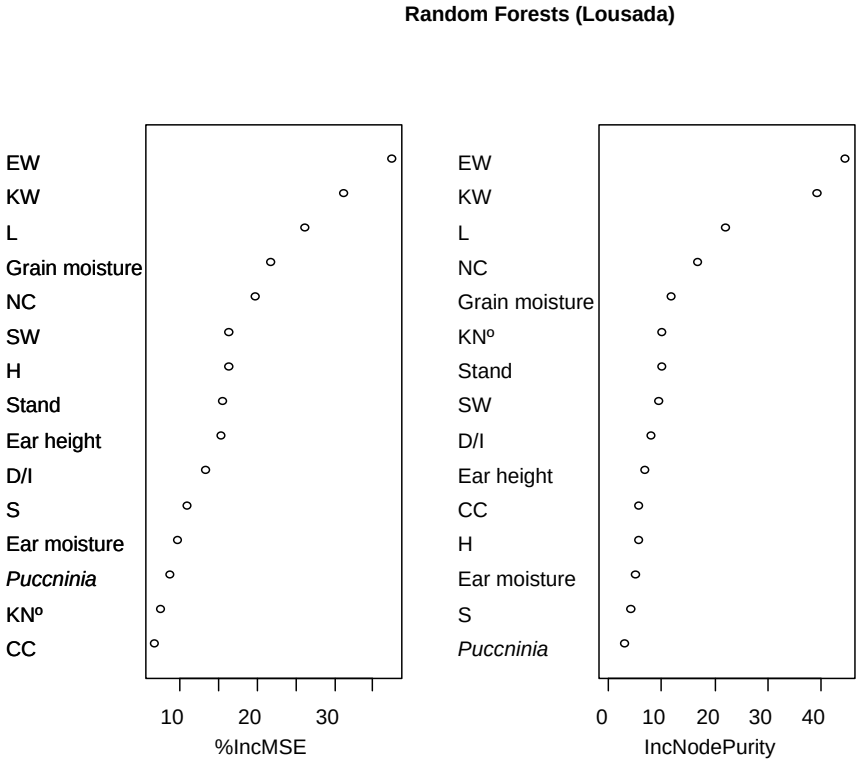

Figure III.4 Mean Decrease Accuracy (% IncMSE) and Mean Decrease MSE (IncNodePurity): there is no clear guidance on which measure to prefer (KUHN et al. 2008). The independent variable is Yield for Lousada. They are presented only the 15 most relevant dependent variables. The percentage of variation explained was 54.4%.

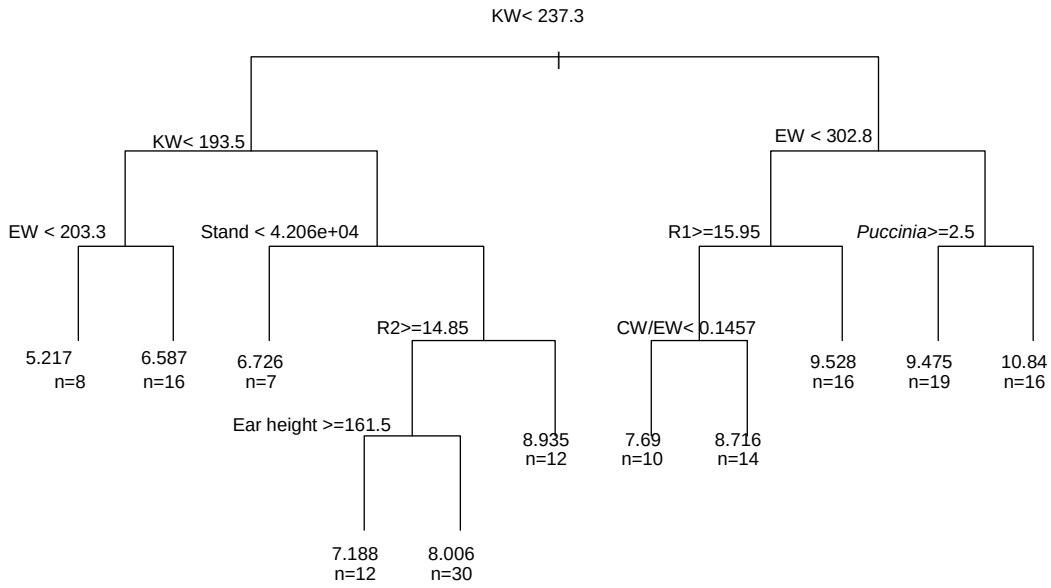

Figure III.5. Decision tree for the independent variable Yield.

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, plant 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, plant stand and thousand kernel weight. The CART analysis revealed that

ear weight and length as well as medulla 2 were used for Lousada (Figure III.6).

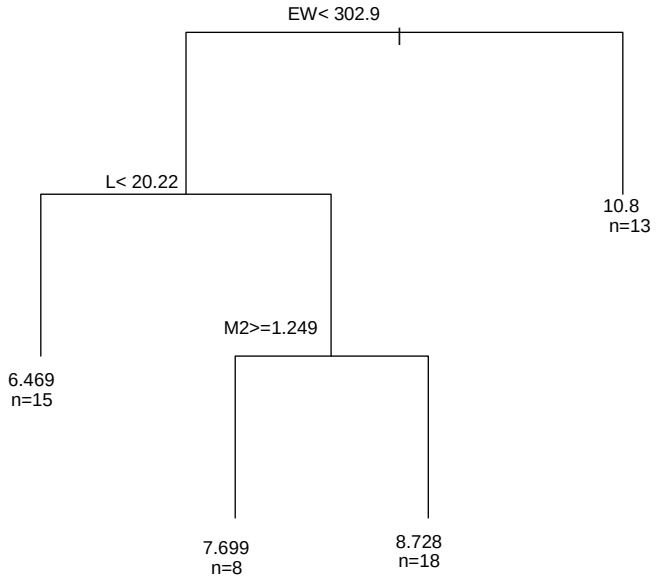

Figure III.6 Decision tree for the independent variable Yield for Lousada.

III.3.2 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 III.2, Table III.4). 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 III.2).

III.4. 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 (Table III.1, Table III.2).

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 grain yield usually occur (Hallauer, 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 Pêgo (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 (Table III.3, Table III.4).

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 (Figure III.1). 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) (Figure III.2).

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 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.

III.5. 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, Pêgo 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.

III.6. Material and Methods

III.6.1 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 Pêgo, 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

The 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 Pêgo's breeding program at ENMP (Elvas Breeding Station). In 1984, with the purpose of evaluating the gene action composition (additive *versus* nonadditive), 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 Pêgo'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, Pêgo 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 (Pêgo, personal communication).

Based on the good results, in 1985 Pêgo 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 (Pêgo, 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.

III.6.2 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) (Table III.6, Figure III.7). The farmer selection pressure ranged from 1 to 5%.

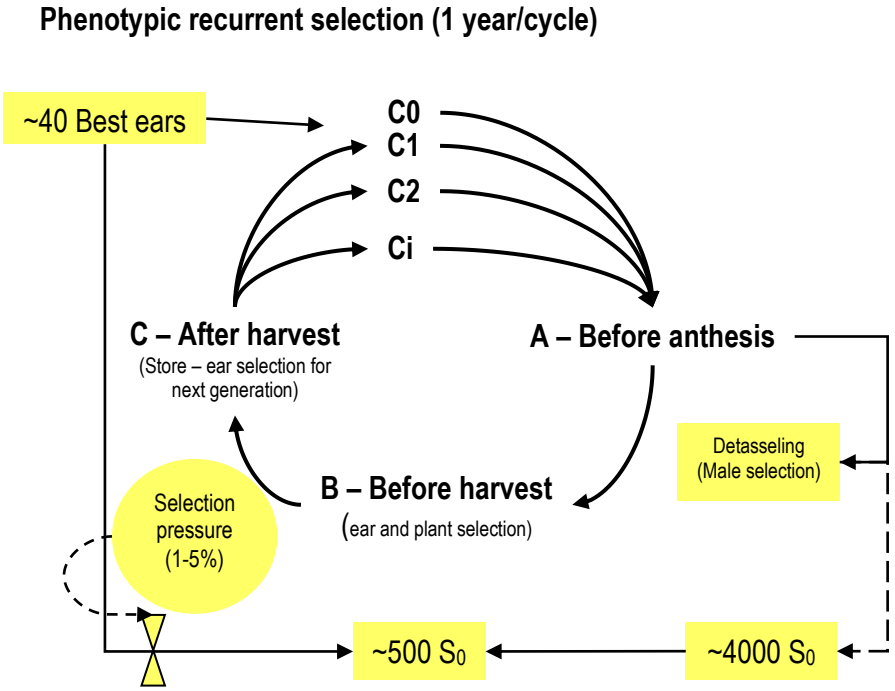

Figure III.7. Phenotypic recurrent selection methodology used in 'Fandango' by the breeder.

Table III.6 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.

III.6.3 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 III.6) 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 III.6):

1. Five to seven cycles of mass selection (breeder 2-3 cycles, farmer 3-4 cycles);

2. 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 Montemoro- 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.

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. 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 plant 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 cobs, to determine the grain weight and the ratio cob/ear weight (Table III.7).

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 III.7, Figure III.8, Figure III.9 (Pêgo, Hallauer 1984; Moreira, Pêgo 2003; Moreira et al. 2008).

Table III.7 Traits measured per location and per plot, codes and respective description

Traits	Measurements	Data/plot			Codes	Scale	
	Iowa	Pt	Plot	Pl or Ears			
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.
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.
Plant 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; 1-4 to pure lines and 5-9 to populations.
Leaf A**n**gle		x	1		N	1 to 9	Angle of the adaxial side of the leaf above the ear with the stalk (5=45º, <5 =<45º and >5 = >45ºC)
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 30º 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 **he**ight, cm		x		20	H		Plant height, from the stalk basis to the last leaf insertion before the tassel

Traits	Measurements Iowa	Data/plot Pt	Plot	Pl or Ears	Codes	Scale	
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	ED1, ED3		Large diameter in the 1/3 bottom and top of the ear respectively;
Ear diameter 2 and 4, cm		x		20	ED 2, ED 4		Small diameter in the 1/3 bottom and top of the ear respectively (90º rotation from large diameter) (cm);
Kernel-row number 1and 2, nº		x		20	R1, R2		Row number in the 1/3 bottom and top of the ear respectively (nº);
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 (35ºC).
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	CD1, 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

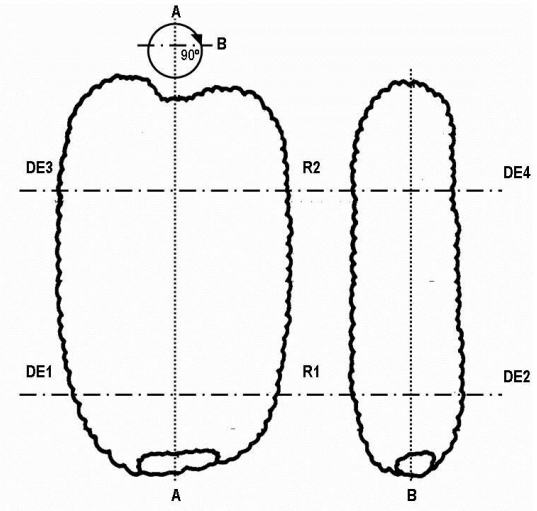

Figure III.8 Two orthogonal views of the same ear showing the way that the two sets of diameters and the two row numbers (R1 and R2) were measured and counted; in position A, the diameters D1 and D3 were measured; in position B (a 90º turn along the length axis), D2 and D4 were measured (Adapted from Pego & Hallauer, 1984)

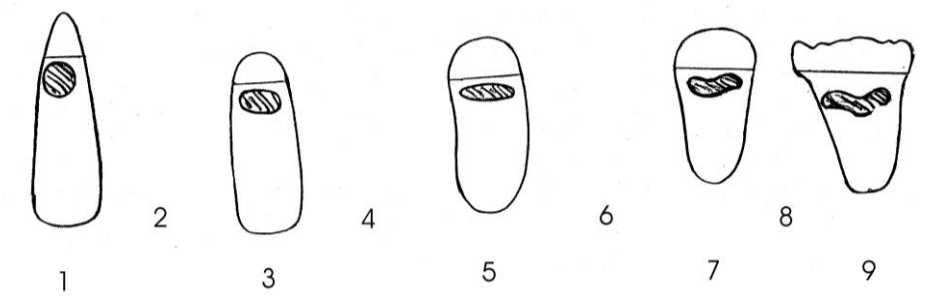

Figure III.9 Fasciation degree (1 – without fasciation and 9 - maximum of fasciation), shape of the ear and from transversal cut view.

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:

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

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, 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.

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 III.7).

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 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'.

III.7. Acknowledgements

Conception and design of the work: ARH, SP, MCVP

Acquisition of data: PMM, JPNS, JPPS

Analysis and interpretation of data: PMM, JMM

Article drafting: PMM, MCVP

Revising it critically: ARH, SP, MCVP, JMM, MM

Populations' development and breeding: SP, MM, EA

III.8. References

Anderson E, (1944) Homologies of the ear and tassel in *Zea mays*. Ann. Missouri Bot. Garden 31: 325-343. doi: 10.2307/2394367

Breiman L (1996) Bagging predictors. Machine Learning 26: 123-140.

Breiman L (2001) Random forests. Machine Learning 45: 5-32.

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and Regression Tree. Chapman and Hall/CRC.

Emerson R, East E. 1913. The Inheritance of Quantitative Characters in Maize, Nebr. Agric. Expt. Sta. Bull 2.

Ferrão JEM (1992) A aventura das plantas e os descobrimentos portugueses. Programa Nacional de Edições Comemorativas dos Descobrimentos Portugueses, Portugal

Friedman JH (1991) Multivariate adaptive regression splines. Ann. Stat. 19: 1–141.

Galinat WC (1980) Indeterminated *versus* determinated years. Maize Genet Coop Newsl 54: 121

Hallauer AR (1992) Recurrent selection in Maize. In: Janick J. (ed) Plant breeding reviews, vol 9. John Wiley & Sons, Inc, pp 115–177

Hallauer AR (1994) Corn genetics and breeding. Encyclopaedia Agric. Sci. 1: 455-467.

Hallauer AR, Carena MJ (2009) Maize breeding. In: Carena, M.J. (Ed.), Handbook of Plant Breeding, Cereals. Springer, New York, pp. xiv, 425.

Hallauer AR, JB Miranda FO (1988) Quantitative genetics in maize breeding, 2nd edn. Iowa State Univ Press, Ames

Hallauer AR, Ross AJ, Lee M (2004) Long-term divergent selection for ear length in maize. In Janick J (ed) Plant breeding reviews, vol 24(2). John Wiley & Sons, Inc, pp 153–168

Hallauer AR, Russel WA, White PR (2000) Registration of BS21(R)C6 and BS22(R)C6 maize germplasm. Crop Sci. 40: 1517.

Hallauer AR, Sears JH (1969) Mass Selection for Yield in Two Varieties of Maize1. Crop Sci 9: 47-50. doi: 10.2135/cropsci1969.0011183X000900010016x

Hastie TR,Tibshirani JH, Friedman (2001) The elements of statistical learning: data mining, inference, and prediction. Springer.

Kuhn J, Egert B, Neumann S, Steinbeck C (2008) Building blocks for automated elucidation of metabolites: Machine learning methods for NMR prediction. BMC Bioinformatics 9: 400

Lilliefors HW (1967) On the kolmogorov-smirnov test for normality with mean and variance unknown. J. Am. Statistical Ass. 62: 399-402

Lonnquist JH (1961) Progress for recurrent selection procedures for improvement of corn populations. Nebrasca Agric. Exp. Stn. Res. Bull. 197.

Mendes Moreira PMR, Pêgo SE, Vaz Patto MC, Hallauer AR (2008) Comparison of selection methods on 'Pigarro', a Portuguese improved maize population with fasciation expression. Euphytica 163: 481-499. doi: 10.1007/s10681-008-9683-8

Moreira PM (2006) Participatory maize breeding in Portugal. A case study. Acta Agronomica Hungarica 54: 431–439. doi: http://dx.doi.org/10.1556/AAgr.54.2006.4.6

Moreira PM, Pêgo S (2003) Pre-breeding evaluation of maize germplasm. The case of a Portuguese open-pollinated variety. In: Abstracts of the Arnel R. Hallauer International Symposium on Plant Breeding. Mexico City, Mexico 17–22 August 2003

Mota M, Bettencourt E, Gusmão L (1978) Cruzamentos múltiplo em milho. In: Relatório das Actividades, Estação Agronómica Nacional. pp. 134-135

Pêgo SE (1982) Genetic potential of Portuguese maize with abnormal ear shape, Ph.D. Thesis, Iowa State Univ.

Pêgo SE, Antunes MP (1997) Resistance or tolerance? Philosophy, may be the answer. In: Proceedings of the XIX – Conference of the International Working Group on Ostrinia. Guimarães Portugal 30th August–5th September 1997

Pêgo SE, Hallauer AR (1984) Portuguese maize germplasm with abnormal ear shape. Maydica 29: 39–53

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.Rproject.org.

Tilman D, Cassman KG, Matson PA, Naylor R, Polasky S (2002) Agricultural sustainability and intensive production practices. Nature 418: 671-677. doi: 10.1038/nature01014

Troyer AF (2000) Origins of modern corn hybrids. Proc Ann Corn Sorghum Res Conf 55: 27–42

Vaz Patto MC, Moreira PM, Carvalho V, Pêgo S (2007) Collecting maize (*Zea mays* L. convar. *mays*) with potential technological ability for bread making in Portugal. Genet Res Crop Evol 54:1555‑1563. doi: 10.1007/s10722-006-9168-3

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. doi: 10.1007/s10681-008-9690-9

CHAPTER IV.

Comparison of selection methods on ‘Pigarro’, a Portuguese improved maize population with fasciation expression

The work presented in this chapter was mostly performed by Pedro Mendes Moreira (see acknowledgments section) and partially included in the following publication: Mendes Moreira, P.M.R., Pêgo, S.E., Vaz Patto, M.C., Hallauer, A.R., 2008. Comparison of selection methods on 'Pigarro', a Portuguese improved maize population with fasciation expression. Euphytica 163, 481-499.

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
/UNL
Knowledge Creation

IV.1. Abstract

In 1984, Pêgo 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). During 20 years of PMB at VASO, mass and S2 recurrent selection were applied on the maize landrace 'Pigarro'. Morphological (*e.g.*, ear length and fasciation level) and yield evaluations were conducted in Portugal (2– 3 locations in 2 years) and in USA (4 locations in one year) using samples from original population, six MS cycles and three S2RS cycles. North American Populations (BS21, BS22, TEPR-EC6) were also included as checks. ANOVA comparisons and regression analyses on the rate of direct response to selection were performed. Response to MS for Iowa showed significant decrease in stalk lodging, while in Portugal ear length significantly decreased, whereas ear diameter, kernel-row number, and fasciation level significantly increased. Selection also significantly increased days-to-silk and anthesis in Portugal. Response to S2 recurrent selection in Portugal significantly increased days-to-silk, uniformity, and cob/ear weight ratio. These results showed that the methods used by farmer and breeder were not effective for significant yield increase, but the ear size increased significantly for mass selection and showed a positive tendency for S2 recurrent selection. Adaptation to farmer needs was maintained for the last cycles of selection.

IV.2. Introduction

IV.2.1 Maize introduction, expansion and genetic adaptation in Portugal

During five centuries, maize shaped the landscape (*e.g.*, terraces, water mills, and store facilities), humans (*e.g.*, traditions, religion, and language), economy (*e.g.*, maize as payment to landlords) and food (*e.g.*, directly for maize bread and indirectly through meat consumption) in Portugal. Hallauer (1994) discussed four distinct stages in maize breeding. The Portuguese maize history begins at the third maize breeding stage, after Columbus' (1492) (Ferrão 1992). The establishment and further expansion of maize during the XVII and XVIII centuries was in the origin of an agricultural revolution, which led to the enhancement of the rural communities' standard of living (Pêgo, Antunes 1997; Moreira 2006). The impact of the maize expansion from the Southern Portuguese region of Algarve till the Northwest of the country led to genetic adaptation to a diversified number of microclimates, according to the sequence of valleys and mountains in these regions.

IV.2.2 Genetic resources and pre-breeding

During the 60's, the diffusion of the hybrid technology has led to a progressive genetic erosion of maize germplasm. In 1975, Portugal took the initiative of a first regional collection of maize germplasm. In

1976, Pêgo proposed and FAO did implement a germplasm bank in Portugal especially devoted to maize.

Several collecting missions were undertaken in the 1970's and 1980's in collaboration with FAO/ IBPGR/IPGRI which formed the basis of the BPGV (Portuguese germplasm bank, with more than 3000 accessions of maize) (Pêgo 1996). One of the most recent collecting missions in Portugal took place in 2005 (Vaz Patto et al. 2007b).

The large amounts of data collected during the evaluation of the accessions following the IPGRI descriptors do not fit the breeder's major needs. A generalized limitation in all germplasm catalogues is the lack of information about inbreeding depression and combining ability, two important traits related with heterosis and yield. This fact reveals the existence of a gap between ''curators'' and ''breeders'' or between ''characterisation'' and ''utilisation''. The same was already stressed by Cooper et al. (2001) that call for the importance of developing pre-breeding methodologies. The overlapping index (Moreira and Pêgo 2003), or the ''HUNTERS'' method (Moreira et al. 2005a, b) and other methods that are being developed by Taba et al. (2003) are good examples of pre-breeding evaluation approaches.

IV.2.3 The VASO project

Agricultural research became associated with farming systems research in the 1980's. Since the late 1980s, participation has become an integrated element of sustainable development strategy.

The critical importance of participation to sustainable development has become widely accepted within the United Nations and among international donor organizations. In the agricultural arena adaptive and farming systems research on agricultural research stations began to include user perspective analyses in the mid-1980s. Substantial work has been done to refine participatory agricultural methodologies, at least within the CGIAR system; *e.g.*, on-farm research methodology (CIMMYT) (Sthapit and Friis-Hansen 2000). Under this context the VASO project was evaluated by Dr. Wayne Haag after its first year and, based on his evaluation, CIMMYT made the decision to completely finance the project.

Altieri and Merrick (1987), Brush (1995, 2000), Bellon (1996), Cooper et al. (2001), Pêgo and Antunes (1997) and Sthapit et al. (2005) have focused on the importance of on-farm conservation as a source of diversity to maintain a dynamic gene flow between germplasm conservation and breeding. Suggested approaches for increasing the diversity available to farmers through participatory varietal selection, participatory plant breeding, collaborative plant breeding and decentralized plant breeding have been used (Cleveland et al. 1999; Ceccarelli et al. 2001; Salazar 2001; Witcombe 2001; Sperling et al. 2001; Machado, Fernandes 2001). In addition to the economic benefits, participatory research has psychological, moral, and ethical benefits, which are the consequence of a progressive empowerment of the farmers' communities. These benefits affect sectors of their life beyond the agricultural aspects by elevating local knowledge to

the role of science (Ceccarelli Grando 2007), and by encouraging interaction between professional plant breeders and other researchers and farmers, with the objective of developing local cropping systems that better meet local needs (Cleveland et al. 1999).

The understanding of the importance of on-farm conservation and participatory plant breeding (PPB) led Silas Pêgo in 1984 to a detailed survey on farmer's maize fields in «Vale do Sousa» Region (VASO) (Pêgo and Antunes 1997; Moreira 2006). VASO was implemented according to an integrant philosophy point of view.

The integrant philosophy approach takes into account not only the agricultural system, but considers the farmer as the most important genetic resource where the decision power resides on (Pêgo, Antunes 1997; Moreira 2006). The goal was to solve the problem of the small Portuguese farmers, 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 achieve this goal, three main decisions had to be taken: (1) the choice of the location to represent the region, (2) the farmer to work with, side-by-side and (3) the germplasm source (Pêgo, Antunes 1997; Moreira 2006). These factors allowed the possibility to test the efficiency of an alternative project to improve the local germplasm in order to be more competitive, at least in certain specific circumstances, in side-by-side comparisons with the local farmers (Pêgo, Antunes 1997).

IV.2.4 Location

The Sousa Valley was chosen because of the following factors: (a) location in a traditional maize area characterized by polycropping systems, where maize still plays an important role; (b) it is one of the most fertile areas of the Northwest region of Portugal; (c) in 1985, 20–25% of its area was planted with hybrids, compared with the 15% national average, creating a perfect situation for developing alternative production systems. It is also in this area where the maize production champion (18 Mg ha^{-1}, with a single cross hybrid) was located; (d) the availability of a basic amount of agro/sociologic/economics data, previously collected by some members of the original multidisciplinary team allowed the breeder a thorough knowledge of the region; (e) the support of a local elite farmers' association (CGAVS) which agreed to be part of the project.

IV.2.5 The farmer

Choosing the right people to work with is also a major decision in an on-farm project, since the system is supposed to work side-by-side with the farmer who will have decision power. All the information gathered was decisive for selecting the farmers. Their initial acceptance and enthusiasm to join the project assured the success of this project. With careful respect for the local traditional agriculture, an agreement was made with the involved farmers. While the breeder would apply his breeding methodologies, the farmers would

continue a parallel programme with their own mass selection criteria. With this tacit agreement between breeder and farmer, three consequences became clear: (1) respecting the ''system'' would imply accepting low input and intercropping characteristics, as well as accepting and respecting the local farmer as the decision maker; (2) with two simultaneous breeding programmes (the farmer's and the breeder's), the farmer would be able to compare the effectiveness of the two systems. This would permit the farmer to base his decisions on solid grounds; and (3) an option for diversity and quality as the first priority trait, due to the choice of local adapted germplasm.

IV.2.6 Germplasm

One of the first aims of VASO project was the selection of a regional open-pollinated variety (OPV), a prerequisite of integrant philosophy option. This selection was done according to 2nd class soils, medium nitrogen inputs, water available, white flint type, bread making characteristics most preferred by the farmers, and its fitness to the traditional polycrop system (with beans and forage). The selected OPV was named by the farmer as 'Pigarro'. 'Pigarro' is of FAO 300 maturity OPV that has white, flint kernels. 'Pigarro' had high levels of root and stalk lodging and was characterized for having high kernel-row numbers (between 18 and 28) because of its strong fasciation expression.

IV.3. Results

IV.3.1 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 found between environments (all locations at Portugal and Iowa) for all traits in the analyses. The genotype x environment interaction (selection cycle x location) was significant for moisture and plant 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.*, plant stand) led to consider Iowa and Portugal as separated groups (analyses not shown).

IV.3.2 Mass selection at Iowa

Significant differences were found among cycles of selection for days-to-silk at Ames. Significant differences were found among environments (field locations) for all traits in the analyses, except for yield. The genotype x environment interaction (selection cycle x field location) was significant for plant stand (Table IV.1).

In the regression analyses were conducted to estimate direct response to selection, and the linear mean squares reveal significant differences for stalk lodging (Table IV.1). Greater proportion of the variation was explained by the linear regression model, providing medium high estimates of response to selection for days-to-silking (78.4%), stalk lodging (82.5%) but low for yield (31.5%) (Table IV.1).

The quadratic regression model for yield also explained a very low amount of variation (37.5%).

Table IV.1 Estimates of linear regression coefficient (b), their standard errors, initial cycle prediction (Ĉ0), coefficients of correlation (R) and % of gain per year (%Gain/Y) for mass selection (20 cycles in Portugal and 15 cycles in Iowa)

Traits	Mass selection Iowa												
	b				Ĉ0	R	%Gain/Y	C	E	Y	CxE	CxY	CxExY
Yield, Mg ha^{-1}	0.009	±	0.0139		2.774483553	0.315	0.33						
Moisture %	0.026	±	0.0568		18.4718192	0.226	0.14		**				
Days-to-silk, nº † (Ames)	0.234	±	0.0925		72.65327381	0.784	0.32	**					
Root lodging %	0.459	±	0.2692		61.30580755	0.649	0.75		**				
Stalk lodging %	-0.763	±	0.2615	*	26.40718917	0.825	-2.89		**				
Stand (Plants ha-1) ‡					59634				**		*		

Traits	Mass selection Portugal												
	b				Ĉ0	R	%Gain/Y	C	E	Y	CxE	CxY	CxExY
Yield, Mg ha^{-1}	0.024	±	0.0196		7.08	0.487	0.35		**	**			
Moisture %	0.060	±	0.0238		28.16	0.748	0.21		**	**	**		**
Days-to-silk, nº †	0.142	±	0.0363	*	61.02	0.868	0.23	**	**	**			
Days-to-silk, nº † end	0.201	±	0.0453	**	65.27	0.893	0.31	**	**	*	*		
Days-to-anthesis, nº †	0.111	±	0.0251	**	58.40	0.893	0.19	**	**	**			**
Days-to-anthesis, nº † end	0.175	±	0.0128	**	62.59	0.987	0.28	**	**	**			
Stand (Plants ha-1) ‡					49963					**			
Overlap index	0.004	±	0.0061		0.34		1.08		*				
Uniformity	0.009	±	0.0057		7.63	0.576	0.12			**			
a**N**gle	0.005	±	0.0077		5.02	0.281	0.10		*			**	
Tassel	0.007	±	0.0073		6.24	0.394	0.11		*	**			
Ear placement	-0.002	±	0.0032		5.40	0.263	-0.04		**	**			
Root lodging %	0.035	±	0.0578		2.53	0.260	1.38			**			**
Stalk lodging %	-0.042	±	0.0575		4.59	0.309	-0.91		**	*			
Plant height, cm ‡	0.433	±	0.3010		225.19	0.541	0.19	**	**		**		
Ear height, cm	0.218	±	0.3241		136.38	0.288	0.16	**	**		**		
Ear Length, cm	-0.055	±	0.0173	*	17.19	0.820	-0.32	**	**		**		
Ear Diameter 1, cm	0.048	±	0.0101	**	5.63	0.904	0.85	**	**				
Ear Diameter 3, cm	0.061	±	0.0104	**	4.59	0.934	1.33	**	**				
Ear Diameter 2, cm	0.035	±	0.0070	**	5.26	0.913	0.67	**	**		**		
Ear Diameter 4, cm	0.032	±	0.0041	**	4.18	0.962	0.77	**	**		**		
Kernel-row number 1, nº	0.264	±	0.0539	**	17.79	0.910	1.48	**	**		*		
Kernel-row number 2, nº	0.266	±	0.0461	**	16.86	0.933	1.58	**	**				

Traits	Mass selection Portugal											
	b			Ĉ0	R	%Gain/Y	C	E	Y	CxE	CxY	CxExY
Fasciation	0.075	±	0.0182	**	1.79	0.881	4.22	**	**				
D/I	-0.009	±	0.0027	*	1.33	0.832	-0.69	**	**		*		
Convulsion	0.034	±	0.0129	*	1.70	0.763	2.01	**	**		**		
Ear weight, g	1.370	±	0.5879		187.83	0.722	0.73	**	**		**		
Kernel weight, g	0.933	±	0.4975		160.50	0.643	0.58	**	**		**		
Cob weight, g	0.437	±	0.1254	*	27.33	0.842	1.60	**	**		**		
Cob/Ear weight	0.111	±	0.0433		14.67	0.753	0.75	**	**				
Ear Moisture %	0.011	±	0.0060		16.45	0.618	0.06	**	**		**		
Kernel dept, cm	0.000	±	0.0006		1.02	0.241	-0.04	*	**				
Kernel number, nº	5.835	±	1.3231	**	456.94	0.892	1.28	**	**		*		
Thousand-kernel weight, g	-1.849	±	0.4173	**	350.44	0.893	-0.53	**	**		**		
Kernel per row, nº	-0.005	±	0.0359		28.71	0.059	-0.02	*	**		*		
Cob diameter 1, cm	0.049	±	0.009	**	3.98	0.924	1.24	**	**				
Cob diameter 3, cm	0.058	±	0.010	**	3.05	0.930	1.89	**	**				
Cob diameter 2, cm	0.031	±	0.006	**	3.48	0.927	0.88	**	**		**		
Cob diameter 4, cm	0.025	±	0.004	**	2.59	0.944	0.97	**	**		**		
Medulla 1, cm	0.027	±	0.0076	*	1.97	0.841	1.35	**	**				
Medulla 2, cm	0.015	±	0.0041	*	1.52	0.850	0.97	**	**		**		
Rachis 1, cm	0.038	±	0.0071	**	3.18	0.921	1.19	**	**				
Rachis 2, cm	0.024	±	0.0049	**	2.65	0.909	0.90	**	**				

* - 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. D/I – determinate and indeterminate ears

%Gain/Y – percentage of gain per year, ANOVA for C-cycles of selection, E-environment; Years; x-interactions; Ĉ - predicted cycle of selection, except for stand that was calculated the average.

For Iowa 5 traits were analysed and for Portugal 43 during 2005 and 13 for 2006

Shaded portions distinguished were Analyses of Variance was not done from the white portions were non-significant differences were registered

IV.3.3 Mass selection at Portugal

Significant differences were found among cycles of selection for end and beginning of silking and anthesis, plant and ear height and all the data related to the ear traits. Significant differences were found among environments (field locations) for all traits in the analyses, except for uniformity and root lodging (Table IV.1). Significant differences were found between year's trial for all traits in the analyses, except for Overlap Index and leaf angle. The genotype

(selection cycle) x environment (field location) interaction was significant for end of days to silk, moisture, plant and ear height, ear length, ear diameter 2 and 4, row number 1, determinate *versus*. indeterminate ears, convulsion, ear, kernel and cob weight, ear moisture, kernel number per ear, thousand-kernel weight, number of kernels per row, cob diameter 2 and 4, and medulla 2 (Table IV.1). Significant differences were found for the genotype x environment x year interaction for moisture, days to anthesis and root lodging. Analysis considering only the Lousada location showed significant differences among cycles of selection for yield, number of days to beginning and ending of silking and anthesis, leaf angle, plant height, ear height, ear length, ear diameters 1, 3, 2 and 4, kernel-row numbers 1 and 2, fasciation, determinate ears, ear, kernel and cob weight, kernel number per ear, ear moisture, thousand-kernel weight, kernel number per row, cob diameter 1, 3, 2 and 4 and medulla 1 and 2, and rachis 1 and 2 (data not shown). The regression analyses were conducted to estimate direct response to selection. The results from the linear mean squares reveal significant differences for beginning and ending of days to silking and anthesis, ear length, ear diameters from 1 to 4, kernel-row number 1–2, fasciation, determinate *versus* indeterminate ears, convulsion, cob weight, kernel number, thousand- kernel weight, cob diameter 1–4, medulla and rachis 1–2 (Table IV.1). A high proportion of the variation was explained by the linear regression model, providing good estimates of response to selection, for moisture (74.8%), days to beginning and ending of silking and anthesis, respectively (86.8%,

89.3%, 89.3%, 98.7%), ear length (82.0%), ear diameter 1, 3, 2 and 4 (90.4%, 93.4%, 91.3% 96.2%), kernel-row number 1 and 2 (91.0% and 93.3%), fasciation (88.1%), determinate *versus* indeterminate ears (83.2%), convulsion (76.3%), ear weight (72.2%), cob weight (84.2%), ratio cob/ear weight (75.3%), kernel number (89.2%), one thousand-kernel weight (89.3%), cob diameter respectively 1, 3, 2 and 4 (92.4%, 93.0%, 92.7% and 94.4%), medulla 1 and 2 (84.1%, 85.0%), rachis 1 and 2 (92.1% and 90.9%). For yield, the linear regression model accounted for 48.7% and the quadratic regression model accounted with 48.8% (Table IV.1).

IV.3.4 Response to S2 recurrent selection

Significant differences ($P< 0.01$ and $P< 0.05$) were found between all locations of Portugal and Iowa for all traits in the analyses. Among selection cycles, significant differences were observed for number of days-to-silk and % root lodging. The genotype x environment interaction was significant for % root lodging.

Significant G x E interaction, plus the different sets of data for Iowa and Portugal and different trial conditions (*e.g.*, plant stand) led to consider Iowa and Portugal as separated groups (data not shown).

IV.3.5 S2 recurrent selection at Iowa

Among selection cycles, significant differences were observed at Iowa, for number of days-to-silk (Ames) and root lodging. Significant

differences were found among environments (field locations) for all traits in the analyses. The genotype x environment interaction was not significant (Table IV.2). Major proportion of the variation was explained by the linear regression model, providing good estimates of response to selection for moisture (75.2%), days-to-silk (84.7%), and stalk lodging % (89.7%). For yield, the linear model accounted for 74.0% of the variation among the several cycles of selection (Table IV.2) and the quadratic model accounted for 98.5% of the variation.

Table IV.2 Estimates of linear regression coefficient, their standard errors, initial cycle prediction, correlation coefficient (R) and % of gain per year for S2 recurrent selection (3 cycles for Portugal and Iowa). Mean traits for standard populations

	Recurrent selection Iowa											Populations Standard Iowa		
	b			Ĉ0	R	%Gain/Y	C	E	Y	CxE	CxY	CxExY	BS21(R)C9	BS22(R)C9	TEPR-EC6
Yield, Mg ha^{-1}	0.040	±	0.0259	2.96	0.740	1.36		**					6.99	6.85	6.44
Moisture %	-0.147	±	0.0914	18.58	0.752	-0.79		**					18.18	20.74	16.62
Days-to-silk, nº † (Ames)	0.175	±	0.0777	73.03	0.847	0.24	**						73.67	73.67	73.67
Root lodging %	-0.705	±	0.7658	62.94	0.545	-1.12	**	**					23.75	30.42	26.90
Stalk lodging %	-0.624	±	0.2177	29.04	0.897	-2.15		**					6.94	9.11	6.27
Stand (Plants ha-1) ‡				59692				**					62960	62304	62741

	Recurrent selection Portugal											Populations Standard Portugal		
	b			Ĉ0	R	%Gain/Y	C	E	Y	CxE	CxY	CxExY	BS21(R)C9	BS22(R)C9	TEPR-EC6
Yield, Mg ha^{-1}	-0.060	±	0.0140	7.22	0.950	-0.84		**					6.84	6.85	7.69
Moisture %	-0.056	±	0.0815	27.64	0.440	-0.20		**	**			**	28.01	26.41	25.28
Days-to-silk, nº †	0.140	±	0.0451	60.76	0.910	0.23	*	**	**				63.56	63.56	62.93
Days-to-silk, nº † end	0.208	±	0.0452 *	64.83	0.956	0.32	*	**					70.44	68.44	67.47
Days-to-anthesis, nº †	0.070	±	0.0404	58.48	0.775	0.12		**	**				62.11	62.33	61.80
Days-to-anthesis, nº † end	0.145	±	0.0343	62.68	0.948	0.23	*	**	**				65.33	65.67	65.80
Stand (Plants ha^{-1}) ‡				50831			*		**				50955	51875	51407
Overlap Index	0.001	±	0.0025	0.40	0.166	0.15						*	0.32	0.42	0.40
Uniformity	0.042	±	0.0065 *	7.60	0.977	0.55	**		**				8.89	8.89	8.53
a**N**gle	0.010	±	0.0141	4.97	0.447	0.20							5.11	4.89	4.33
Tassel	0.010	±	0.0151	6.21	0.424	0.16			**				4.11	4.11	4.87

	Recurrent selection Portugal									Populations Standard Portugal		
	b	Ĉ0	R	%Gain/Y	C	E	Y	CxE	CxY	CxExY	BS21(R)C9	BS22(R)C9TE	PR-EC6
Ear placement	-0.017± 0.0158	5.300	.598	-0.31		**	**				4.33	4.22	4.33
Root lodging %	-0.015± 0.0302	2.140	.340	-0.72			**			**	0.00	0.00	0.01
Stalk lodging %	-0.066± 0.1958	3.880	.231	-1.70	**	**	**	**			0.00	0.00	0.01
Plant height, cm	-0.603± 0.3033	223.110	.815	-0.27	**	**		**			216.18	210.46	199.86
Ear height, cm	-0.493± 0.2179	132.730	.848	-0.37	**	**		**			109.15	96.67	99.54
Ear Length, cm	0.060± 0.0470	17.050	.671	0.35	**	**		*			14.62	16.81	14.79
Ear Diameter 1, cm	0.005± 0.0077	5.500	.408	0.09		**		*			4.73	4.73	4.43
Ear Diameter 3, cm	-0.010± 0.0058	4.500	.781	-0.23		**					4.23	4.21	3.96
Ear Diameter 2, cm	0.000± 0.0080	5.180	.010	0.00	*	**		**			4.62	4.63	4.34
Ear Diameter 4, cm	-0.007± 0.0077	4.170	.534	-0.17	**	**		**			4.11	4.12	3.90
Kernel-row number 1, nº	-0.045± 0.0295	17.240	.735	-0.26		**					16.24	14.80	14.80
Kernel-row number 2, nº	-0.086± 0.0514	16.520	.764	-0.52	**	**					15.18	14.57	14.44
Fasciation	-0.025± 0.0156	1.730	.753	-1.46	**	**		**			1.11	1.07	1.07
D/I	-0.018± 0.3758	*1.385	60.957	-1.26	**	*		**			1.06	1.02	1.16
Convulsion	0.000± 0.0198	1.750	.015	-0.02	**	**		**			1.71	1.46	1.41
Ear weight, g	0.179± 0.7256	178.140	.171	0.10	**	**		**			156.52	172.89	147.72
Kernel weight, g	-0.178± 0.5907	152.140	.208	-0.12	*	**		**			135.56	148.32	126.85
Cob weight, g	0.356± 0.1384	26.000	.876	1.37	**	**		**			20.96	24.57	20.86
Cob/Ear weight	0.185± 0.0044**	14.690	.999	1.26	**	**					13.51	14.46	14.35
Ear Moisture %	-0.022± 0.0107	16.480	.817	-0.13	**	**		**			15.69	15.75	15.55
Kernel dept, cm	-0.002± 0.0004 *	1.000	.966	-0.22	*	**		*			1.19	1.10	1.15
Kernel number, nº	-0.405± 1.7034	434.700	.166	-0.09	*	**					417.57	453.82	427.72
Thousand-kernel weight, g	-0.399± 1.0895	351.730	.251	-0.11	**	**		**			327.01	326.02	296.48
Kernel per row, nº	0.007± 0.0996	28.480	.046	0.02	**	**		*			28.67	32.38	30.88
Cob diameter 1, cm	0.014± 0.0072	3.910	.811	0.36	**	**					5.80	6.33	5.50
Cob diameter 3, cm	-0.002± 0.0059	2.980	.210	-0.06		**		*			4.40	4.86	4.57
Cob diameter 2, cm	0.012± 0.0097	3.470	.660	0.35	**	**					5.59	6.10	5.22
Cob diameter 4, cm	0.000± 0.0064	2.650	.026	-0.01	*	**		**			4.24	4.76	4.45
Medulla 1, cm	0.017± 0.0065	1.890	.880	0.90	**	**		**			0.94	1.15	0.96
Medulla 2, cm	0.011± 0.0050	1.500	.844	0.74	**	**		**			0.78	0.98	0.81
Rachis 1, cm	0.015± 0.0043	3.110	.926	0.47	**	**					2.02	2.16	1.93
Rachis 2, cm	0.013± 0.0095	2.630	.705	0.51	**	**					1.80	1.93	1.73

* Significant at 0.05 probability levels; ** Highly significant at 0.01 probability levels; a Number of days from date of planting to date of flowering; b The plant stand correspond to the average of the correspondent cycles. D/I—determinate and indeterminate ears

%Gain/Y—percentage of gain per year, ANOVA for C-cycles of selection, E-environment; Years; 9-interactions; Ĉ -—predicted cycle of selection, except for plant stand that was calculated the average

For Iowa 5 traits were analysed and for Portugal 43 during 2005 and 13 for 2006. Populations standard and respective mean for each trait shaded portions distinguished were Analyses of Variance was not done from the white portions were non-significant differences were registered

IV.3.6 S2 recurrent selection at Portugal

Among selection cycles, significant differences were observed in number of days to beginning and ending of silking and ending of anthesis, plant stand, uniformity, stalk lodging, plant and ear height, ear length, ear diameter 2 and 4, row number 2, fasciation, determinate *versus* indeterminate, ear convulsion, ear, kernel and cob weight, ratio cob and ear weight, ear moisture, kernel depth, kernel number, thousand-kernel weight, number of kernels per row, cob diameter 1, 2 and 4, medulla 1 and 2, and rachis 1 and 2. Significant differences were found among environments for all traits in the analyses, except for plant stand, overlapping index, uniformity, leaf angle, tassel and root lodging. Significant differences were found between year's trials for all traits in the analyses, except for yield end of silking, overlapping index and leaf angle. The genotype x environment interaction was significant for stalk lodging, plant and ear height, ear length, ear diameter 1, 2 and 4, fasciation, determinate *versus* indeterminate ears, convulsion, ear, kernel and cob weight, ear moisture, kernel depth, thousand-kernel weight, number of kernels per row, cob diameter 3 and 4, and medulla 1 and 2 (Table IV.2). The genotype x environment x year interaction was significant for moisture, overlapping index and % root lodging.

Analysis considering only Lousada showed significant differences among cycles of selection for number of days to beginning and ending of silking and anthesis, ear location, ear length, ear diameter 4, convulsion, ear and cob weight, ear moisture, cob diameter 1 and 2, medulla 1 and 2, and rachis1 and 2.

In the regression analyses conducted to estimate direct response to selection, the results from the linear mean squares reveal significant differences for end of silking, uniformity, determinate *versus* indeterminate ears, ratio cob/ear weight, and kernel depth (Table IV.2). Major proportion of the variation was explained by the linear regression model, providing good estimates of response to selection for beginning and end of silking and anthesis (91.0%, 95.6%, 77.5%, 94.8%), uniformity (97.7%), plant and ear height (81.5% and 84.8%), ear diameter 3 (78.1%), kernel-row number 1 and 2 (73.5% and 76.4%), fasciation (75.3%), determinate *versus* indeterminate ears (95.7%), cob weight (87.6%), ratio cob/ear weight (99.9%), kernel depth (96.6%), cob diameter 1 (81.1%), medulla 1 and 2 (88.0% and 84.4%), rachis 1 and 2 (92.6% and 70.5%) and yield (95.0%) of the variation among the several cycles of selection. For yield, the quadratic model had the same result than for linear regression model.

IV.3.7 Standard North American populations

The standard populations showed no significant differences between Iowa and Portugal, which did not happen with 'Pigarro' cycles presenting a variation between Iowa and Portugal of 144% for mass selection and 155% for S2 recurrent selection. These results can be caused by the non adaptation of 'Pigarro' to Iowa locations, but also to mechanical harvest used in Iowa (high root and stalk lodging) (Table IV.2).

IV.4. Discussion

IV.4.1 Response to mass selection

Response to mass selection for Iowa revealed a significant decrease of stalk lodging indicating a positive response for higher densities (Table IV.1). Response to mass selection in Portugal, reveal significant increase in days to silking and anthesis (Table IV.1), which is related with an increase of lateness. Hallauer and Miranda (1988) reveal that during mass selection there is a tendency for latter material. During mass selection for larger ears, days to- anthesis and to silking significantly increased. Both traits are related positively with yield (Hallauer and Miranda 1988). Ear length significantly decreased but ear and cob diameter and number-of rows significantly increased because of increased expression of fasciation, which agrees with reports by Hallauer and Miranda (1988) and Pêgo (1982). Similar outcomes were observed in long-term divergent selection for ear length in maize (Hallauer 1992) and also in reports by Emerson and East (1913) between ear length and number of rows and between diameter and kernel-rows and seed size. The determinate *versus* indeterminate ears significantly increased. Galinat (1980) indicated that indeterminate ears may elongate under unusually favorable conditions, and kernel row arrangement became more irregular (convulsion), which could be related with fasciation (Table IV.1).

Results from Vaz Patto et al. (2007a) using 16 SSR on 3 selection cycles (C0-84, C9-93, C20-04) of 'Pigarro' revealed that no effective

loss of genetic diversity had occurred during the selective adaptation to the farmer's needs and the regional growing conditions. Variation among selection cycles represented only 7% of the total molecular variation indicating that a great proportion of the genetic diversity is maintained in each selection cycle. Genetic diversity has not been reduced from the 'Pigarro' breed before 1984 to those improved after 2004, but the genetic diversity maintained is not exactly the same. These ''qualitative'' changes also may have a phenotypic expression since as described in the present work, the evaluation of phenotype of individual plants, revealed an increase in ear size under mass selection. A few of the SSR molecular markers, used on the work of Vaz Patto et al. (2007a), exceeded the expected Hardy-Weinberg equilibrium and mapped in the maize genome at locations (bins) where Quantitative Trait *Loci* (QTL) related to yield were described by others (Vaz Patto et al. 2007a). As pointed out by Butrón et al. (2005), the directional selection observed on these SSR markers could suggest the presence of QTLs controlling the real selection trait or traits linked to these markers. Based on the present results, the identification of the genetic control of the detected decrease in ear length and increases in ear diameter, kernel-row number, and cob diameter could start by looking for significant associations with the respective molecular markers.

IV.4.2 Response to S2 recurrent selection

No significant responses to S2 recurrent selection in Iowa were observed (Table IV.2). The response to S2 recurrent selection in Portugal increased days-to-silk that is positively and significantly related with yield (Hallauer, Miranda 1988) and uniformity of plants. In the case of ear traits, ratio cob/ear weight significantly increased and kernel depth significantly decreased.

Contrary to what happened with mass selection here, ear length increased, ear and cob diameter and number-of-rows decreased; similarly to what is described by Lopez-Reynoso and Hallauer (1998) and Hallauer et al. (2004). The determinate *versus* indeterminate ears also significantly increased. It was observed also a tendency of increased kernel number, but at the same time the one thousand-kernel weight decreased, which means that there are smaller kernels per ear. A significant response to selection for cob/ear weight ratio indicates a significant increase of cob weight, and in parallel a decrease in fasciation expression, which is strongly correlated with low cob/ear weight, because fasciated ears are generally hollow. Kernel depth significantly decreased, which contributed to the yield decrease (Table IV.2).

The negative tendency observed for yield under S2 recurrent selection was not totally unexpected because of limited number of cycles of selection (three cycles in this case), but 95% of the variation was explained by the linear regression model. In case of Lousada (where 'Pigarro' was selected) the response had a better fit to a

quadratic model (78.85%) than to the linear regression model (55.14%). The difference in response could be related to the loss of diversity, but molecular studies are under way to determine if changes in diversity have occurred. Other traits such as uniformity, did register an increase during selection. Among ear traits, ratio cob and ear weight increased, kernel depth decreased, and kernel depth was positively correlated with kernel-row number but at the expense of kernel size, which agrees with the results here obtained (Table IV.2).

IV.4.3 Comparison of selection methods

The lack of significant progress in yield for both selection methods could be 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 yield and stalk lodging, the exclusion of stalk lodged plants reduces the intensity of selection for yield from 7.5 to 27.4%. On the other hand trials were done in mixtures of late material in 2005 and early to late materials in 2006, which could affect the pollen flow in plots trials and the yield potential for each plot.

Despite the absence of significant yield progress for both methods, mass selection in Portugal was positively effective for increase ear size (significant differences for ear, cob, medulla and rachis diameters, kernel-row number, fasciation and convulsion), but had a negative effect to increase significantly the number of days to begin

and end of silking and anthesis and significantly increasing cob weight. In Iowa stalk lodging significantly decreased, which indicates better stalk resistance as response to selection.

Results from both selection methods used in VASO project suggest mass selection as better than S2 recurrent selection due to the following reasons: (A) Mass selection: had a slight increase in yield, and is a cheaper methodology, technically more accessible to farmers. One cycle of selection can be completed each summer session, and conservation *in situ*/on-farm of diversity is effective (Vaz Patto et al. 2007a); one disadvantage was the lack of significant yield increase, which could be a disadvantage if this maize was not orientated for human food niche market. (B) S2 recurrent selection has the advantages of being more adapted for a breeding programme on station due to its uniformity, almost absence of fasciation *i.e.*, less variation in ear diameter and greater reduction of root and stalk lodging %. However it is a more complex, more time consuming approach (4 seasons to complete one cycle of selection). Nevertheless more cycles of recurrent selection would be needed to check if tendency of yield decreasing is due to lost of diversity or due to selection procedures for stalk and root lodging %, which could affect yield.

The anthropological and sociological objective of participatory breeding needs: (1) Learning more about how plant breeding on farms changes plant breeding itself, for example is ''on-farm'' plant breeding simply conventional plant breeding onfarms, or is it a whole

different kind of plant breeding for the future? (Powell 2000); (2) the definition of ‘‘yield’’ needs to be broadened to include the total yield of the farm not just a single crop (Pêgo, Antunes 1997; Powell 2000); (3) It is important for breeders to work with other people involved in the food production ‘‘chain’’ like traditional grain millers and also bakers (Powell 2002). During VASO project, farmer had the chance to compare breeding methodologies side by side with breeder, *i.e.*, his decisions were based on his live experience (Pêgo, Antunes 1997). The selection for big ears led to the winning of several trophies by the farmer at ‘‘Sousa Valley Best Ear’’. This contest allows the recognition of the farmer by the community, but also attracts new farmers and germplasm for new initiatives.

Compared with the literature on collaborative plant breeding, VASO can be considered exemplary in regards to duration. Hence this project faces the problem of diminishment of smallholder farming as a viable way of life in Portugal and the socio-economic ‘‘pull’’ factors that remove younger generations from the farm (Powell 2000; Vaz Patto et al. 2007b).

IV.5. Conclusions

The results from response to mass selection in Portugal revealed that ear length significantly decreased and simultaneously, ear diameter, kernel row number and fasciation significantly increased. This selection also led to significant increases of days-to-silk and anthesis. In the case of the Response to S2 recurrent selection in Portugal, data

analyses revealed that days-to-silk, uniformity, ratio cob ear/ ear weight, significantly increased.

These results showed that the methods used by the farmer (mass selection with 1–5% of selection pressure) and by the breeder (S2 recurrent selection with 15–20% of selection pressure), were not sufficient for significant yield increase. The last cycle of both selection methods maintain the ability for polycrop systems and quality for bread production (empirically tested), but no study was done yet to compare differences on these traits along the cycles of each selection method. Maize bread making quality quantification parameters have just been defined (personal communication, Carla Brites), and will be used for future comparisons.

The lack of significant progress in yield for 'Pigarro' suggests new experiments: plant densities trials with maize only or in polycrop system, fertilization levels trials, continuation of the studies of overlapping index (Moreira and Pêgo 2003), but also approaches such as doubled haploid lines or synthetic populations.

Hybrid populations' development could contribute also to yield progress and to avoid the collapse of some interesting germplasm. This strategy can be applicable if farmers associations for specialties (*e.g.*, bread maize) are willing to pay to the farmers (onfarm conservation of populations and hybrid populations seed) and to the breeder (*e.g.*, breeding plan, monitoring). This means that apparently contradictory integrant and productivist philosophies have their specific niches of application and some ''hybrid'' philosophic

adaptations will be preferred in certain situations. This scenario is supported by enthusiastic results from hybrid populations (Silas Pêgo personal communication).

The PPB can improve populations in conservation *in situ*/on-farm strategy that could help to design better synthetic populations or hybrids for low input and organic agriculture. Besides Pigarro, other landraces are being evaluated, under the same prebreeding work, for PPB implementation. This PPB programs should be planned as rural development strategies, where specialties and traditional food are the major output, but where hybrid industry can search for germplasm that is being produced in a coevolutionary process and in a low input or organic system.

Previous knowledge on molecular diversity evolution through PPB highlights 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) will be subject of analysis in an ongoing QTL analysis study.

IV.6. Material and Methods

IV.6.1 Germplasm selection

Under the VASO project two simultaneous experiments were conducted to include both the yield component and pest and diseases performance. The breeding approach was conducted based on the concepts of quantitative genetics in population improvement for two main recurrent selection methodologies: phenotypic recurrent selection and S2 lines recurrent selection (Pêgo, Antunes 1997).

IV.6.2 Phenotypic recurrent selection (mass selection)

The phenotypic recurrent selection or mass selection (since 1984 till present), program included two parental controls (stratified mass selection with parental control c = 1.0). This is an improved extension of the mass selection procedure commonly used by farmers (for one parental control c = 0.5). The farmer was advised to conduct selection under a three step sequence (A–B–C). The first two steps (A and B) in the field and the third one (C) at the storage facilities (Figure IV.1, Table IV.3):

(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 plants 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 finally shelled and mixed together to form the next generation seed. The farmer selection pressure ranged from 1 to 5%.

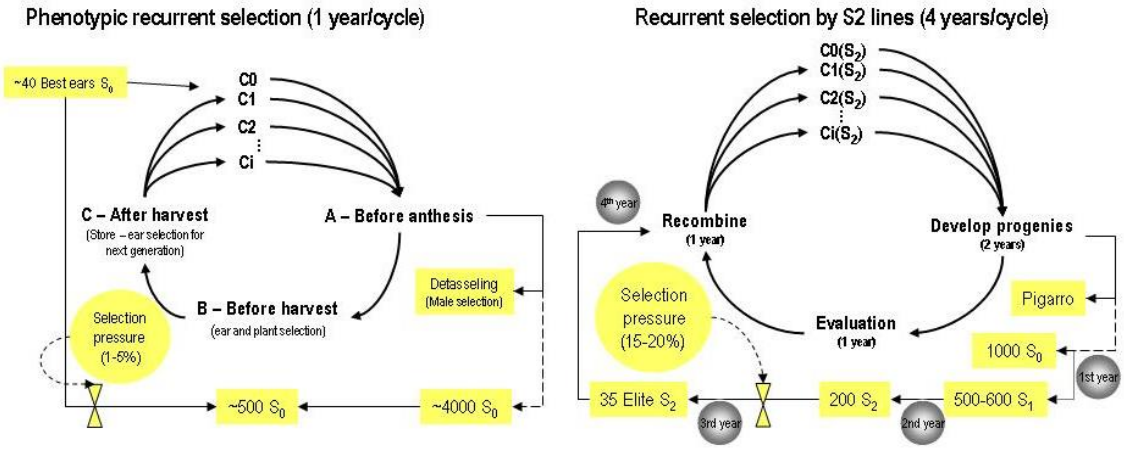

Figure IV.1 Phenotypic recurrent selection and recurrent selection by S2 lines methodologies.

Table IV.3 Breeding methodologies applied to 'Pigarro' since 1984, selected cycles for trials, evaluation (locations and years), seasons per cycle and standard populations used.

Selection method																						**Year.Cycle-1**	
Year	1984	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	00	01	02	03	2004		
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	1	
Recurrent Selection (S_2)						C1(S2)					C2(S2)				C3(S2)							4	
Evaluation trials																						**Standard populations**	
Mass Selection	C0-84				C4-88		C6-90			C9-93			C12-96			C15-99					C20-04	BS21(R)C9	
Recurrent Selection (S_2)						C1(S2)-89					C2(S2)-94				C3(S2)-98							BS22(R)C9	TEPR-EC6
Locations (with 3 replications)																							
Iowa (2005)	4				4	4	4			4	4		4		4	4					**	4	4
Portugal (2005)	3				3	3	3			3	3		3		3	3					3	3	3
Portugal (2006*)	2				2	2	2			2	2		2		2	2					2		2
Multiplication seed stock	05				05	05	05			05	05		05		05	05						05	05

* - drought after sowing at Montemor-o-Velho location lead to data exclusion; ** - C20-04, due to seed problems during germination, data were excluded; Cx(S2)-y, where C-cycle, x-number of cycles, S2 – if selection by S2 lines, y – year correspondent to cycle of selection

IV.6.3 Recurrent selection by S2 lines

S2 recurrent selection was applied because it takes into consideration the additive component of the genetic variance ($3/2\sigma_a^2$) (Hallauer 1992). Selection was organized in a four season scheme three cycles completed (Figure IV.1, Table IV.3).

Season (1) 1000 S0 plants were selected and selfed, from which 500–600 S1's were selected at harvest;

Season (2) 500–600 S1's were planted and selfed to obtain the S2 seed and at harvest the best 200 ears were selected;

Season (3) the selected S2's were submitted to a yield trial in a randomized complete block design and tested for yield performance, pest and disease tolerance, and stalk quality; and

Season (4) using remnant S2 seed, the best 30–35 S2 lines (15–20%, selection pressure) were planted and recombined through controlled pollination to form the first cycle C1(S2) seed. The same sequence was conducted until the third cycle C3(S2) was completed.

IV.6.4 Germplasm evaluation

IV.6.4.1 Germplasm management

Since the beginning of the VASO Project, phenotypic data were collected and seed of each selection cycle of 'Pigarro', either from phenotypic recurrent selection or from S2 recurrent selection, was kept at 4ºC in NUMI

(Maize Breeding Station, Braga, Portugal) cold storage facilities. Seed of cycles C0-84, C4-88, C6-90, C9-93, C12- 96, C15-99, C20-04 of phenotypic recurrent selection and, C1(S2)-89, C2(S2)-94, C3(S2)-98 of S2 recurrent selection, from NUMI (Table IV.3) were chosen and used for trials in 2005. In parallel, the same selection cycles were multiplied by hand pollination to increase the seed stock (*e.g.*, seed for 2006 trials). All pollinated ears were harvested, dried at approximately 35ºC to uniform moisture level of 13–14%.

IV.6.4.2 Evaluation trials

To determine the effectiveness of both methods, trials were set with (Table IV.3):

1. Original population C0-84 was used in both methods;

2. Mass selection included 6 cycles and S2 recurrent selection include 3 cycles;

3. Three replication trials were used for each entry and location;

4. Trials were conducted in four locations at Iowa- USA (Calumet, Kanawha, Ames and Nashua) during 2005 and three locations at Portugal during 2005 (Lousada, Montemor-o-Velho, and Coimbra) and two locations in 2006 because the Montemor-o-Velho location was lost due to unexpected drought after sowing.

The North American populations BS21(R)C9, BS22(R)C9 (Hallauer et al. 2000), were included on 2005 trials and TEPR-EC6 (Troyer 2000) was

included on 2005–06 trials. These populations were included as standards regarding the cycle of Pigarro. They were used to better understand the differences between USA and Portugal environments, and because these populations are better known than 'Pigarro' by the scientific community.

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 mechanical and/or hand weeded as necessary.

IV.6.4.3 Data collection

Data were obtained in all the field trials for final plant 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 cobs, in order to determine the grain weight and the ratio cob/ear weight (Table IV.4).

In the particular case of Portugal trials plant and ear heights were recorded using 20 random plants per plot. After harvest, 20 random ears of each plot were dried at 35ºC to approximately 15% grain moisture. Ear data included overlapping index, ear length, ear diameters, kernel-row number and fasciation and other traits that are summarized on Table IV.4, Figure IV.2 and Figure IV.3 (Pêgo and Hallauer 1984; Moreira and Pêgo 2003; Moreira et al. 2005a, b; Vaz Patto et al. 2007b).

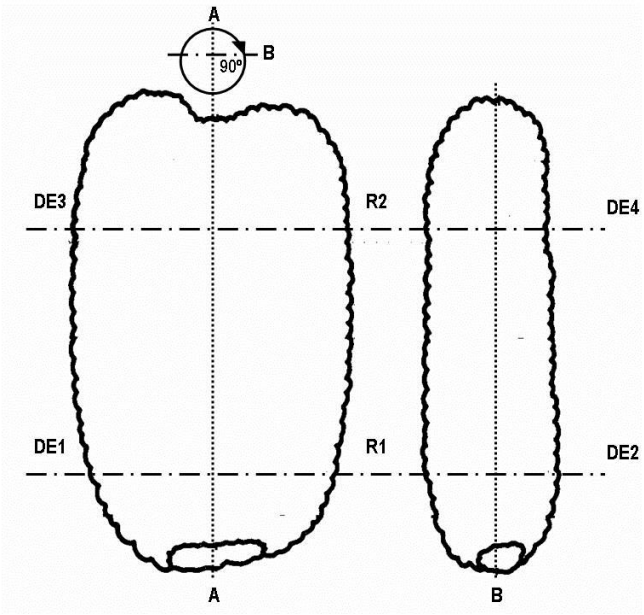

Figure IV.2 Two orthogonal views of the same ear showing the way that the two sets of diameters and the two row numbers (R1 and R2) were measured and counted; in position A, the diameters D1 and D3 were measured; in position B (a 90º turn along the length axis), D2 and D4 were measured (Adapted from Pêgo & Hallauer, 1984)

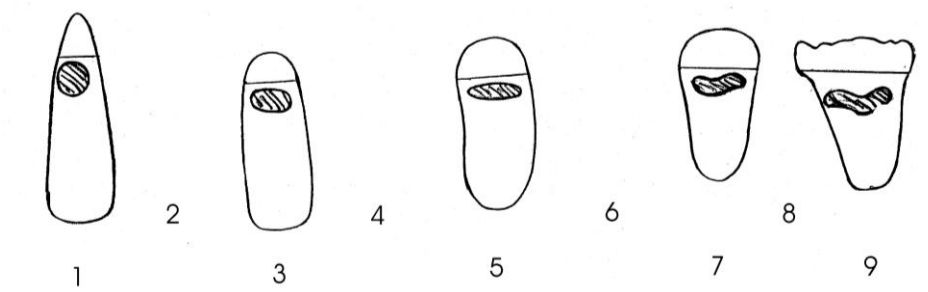

Figure IV.3 Fasciation degree (1 – without fasciation and 9 as a maximum of fasciation), shape of the ear and from transversal cut view.

The overlapping index determination enables the knowledge of a population concerning the relative amount of theoretical allogamy *versus* autogamy. The theoretical reasoning employed consists in assuming that all the polinization occurs only under gravity influence, so that when maize plant has the hypothesis of flowering overlapping, this selfing probability 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 are used in the mathematical expression as follow (Equation IV.1):

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

Equation IV.1

Table IV.4 Traits measured per location and per plot, codes and respective description

Traits	Measurements			Data/plot		Codes	Scale	
	Iowa	Pt05	Pt06	Plot	Pl or Ears			
Yield, Mg ha^{-1}	x	x	x	1		Yield		Grain yield (Mg ha-1) 15% moisture, 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)
Moisture %	x	x	x	1		Moisture %		Moisture content, 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	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	x	1		Ff		The end of days-to silk (from planting until 50% of the plants in the plot begin and finish silk emergence.
Days-to-anthesis, nº †		x	x	1		Mi		The beginning of days-to anthesis, i.e., from planting until 50% of the plants in the plot begin anthesis
Days-to-anthesis, nº † end		x	x	1		Mf		The end of days-to anthesis (from planting until 50% of the plants in the plot end silk emergence
Stand	x	x	x	1		Plants ha-1		Thousands of plants per hectare
OI		x	x	1		OI		**O**verlap **I**ndex, This method enables the knowledge of a population concerning the relative amount of theoretical allogamy versus autogamy
U		x	x	1		U	1 to 9	Uniformity, (1-minimum uniformity and 9 – maximum) 1-4 to pure lines and 5-9 to populations.
N		x	x	1		N	1 to 9	Angle of the adaxial side of the leaf above the ear with the stalk (5=45º, <5 =<45º and >5 = >5=45ºC)
T		x	x	1		T	1 to 9	Tassel branching. 1- Absent tassel (Inbreeds and hybrids) 9- a much branched tassel (frequent in populations with abnormal fasciated ears).
E		x	x	1		E	1 to 9	Ear placement, 5- indicates that the ear is located in the middle of the plant.
Root lodging %	x	x	x	1		R	%	Root lodging (percentage of plants leaning more than 30º from vertical
Stalk lodging %	x	x	x	1		S	%	Stalk lodging (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.
Plant height, cm		x			20	Plant height		Plant height, from the stalk basis to the last leaf insertion before the tassel

Traits	Measurements			Data/plot		Codes	Scale	
	Iowa	Pt05	Pt06	Plot	Pl or Ears			
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	ED1, ED3		Large diameter in the 1/3 bottom and top of the ear respectively;
Ear Diameter 2, cm		x			20	ED2, ED4		Small diameter in the 1/3 bottom and top of the ear respectively (90º rotation) (cm);
Kernel-row number 1, nº		x			20	R1, R2		Row number in the 1/3 bottom and top of the ear respectively (nº);
Fasciation		x			20	Fa	1 to 9	Fasciation degree (1 – without fasciation and 9 as a maximum of fasciation) measures
Determinate versus indeterminate ears		x			20	D/I		Top of the ear full of grain, determinate (2) or not indeterminate (1) ear, average value is calculated.
Convulsion		x			20	CV	0 to 5	Convulsion intensity, kernel-row arrangement in the ear (0 - without convulsion, regular kernel-row arrangement, 5 – maximum of convulsion, without kernel-row arrangement)
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, g		x			20	CW		Cob weight, adjusted to 15% moisture
Cob/ear weight		x			20	CW/EW		Ratio cob/ear weight , indicates the percentage of cob weight in the ear weight
Ear Moisture %		x			20	Ear Moisture %		Determination of % moisture content for ears submitted to 35ºC after harvest.
Kernel dept, cm		x			20	KD		Kernel dept, 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-kernel weight at 15% moisture content
Kernel per row, nº		x			20	KR		Kernel number per row
Cob diameter 1, 3, 2 and 4 cm		x			20	CD1, 3, 2 and 4		CD1, **3**, **2** and **4** measure de same way for DE's
Medulla 1 and 2, cm		x			20	M1, M2		Large and small length of medulla respectively, cob is cut in the Diameter 1 position (IBPGR, 1991; IPGRI, 2000)
Rachis 1 and 2, cm		x			20	Rq1, Rq2		Large and small length of rachis, cob is cut in the Diameter 1 position (IBPGR, 1991; IPGRI, 2000)

This formula allows one to elaborate a reasoning, under its own limitations, such as: OI is limited to 1 (100%); OI is either positive (some overlapping) or negative (overlapping does not occur).

IV.6.4.4 Data analyses

All experiments were analysed as randomized complete block designs, with three replications. Analyses of variance were calculated in each selection method for all environments (locations) in combination with years (Iowa 05; Portugal 05–06). The same analyses were performed for 2 subgroups of each selection method based on Iowa and Portuguese locations. When significant differences were detected, post-hoc comparisons with Sheffe test were performed.

A regression analysis was conducted separately for each selection method, both for Portuguese and Iowa locations. In response to mass selection, regression analyses included 15 cycles for Iowa and 20 cycles for Portugal. Three cycles of recurrent selection were evaluated in Iowa and Portugal.

Response to selection for several traits was evaluated for each of the four subgroups (Iowa and Portugal both with mass and recurrent selection) 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 in a year bases).

For some traits, such as yield, quadratic regression model was also used when significant deviations from the linear regression model were detected. Note that for Iowa locations or Iowa plus Portugal locations, only 15 cycles of mass selection were analysed due to C20-04 exclusion. The C20-04 was excluded, due to poor germination. Number of days-to-silk was considered only at Ames.

Based on Table VII.1, Table VII.2comparisons between both methods of selection were done.

IV.7. Acknowledgements

Conception and design of the work: ARH, SP, MCVP

Acquisition of data: PMM

Analysis and interpretation of data: PMM

Article drafting: PMM, MCVP

Revising it critically: ARH, SP, MCVP

Population development and breeding: SP, farmer FM

IV.8. References

Altieri MA, Merrick LC (1987) *In situ* conservation of crop genetic resources through maintenance of traditional farming systems. Econ Bot 41: 86–96. doi: 10.1007/BF02859354

Bellon MR (1996) The dynamics of crop intraspecific diversity: a conceptual framework at the farmer level. Econ Bot 50:29–36

Brush SB (1995) *In situ* conservation of landraces in centres of crop diversity. Crop Sci 35: 346–354. doi:10.2135/cropsci1995.0011183X003500020009x

Brush SB (2000) The issues of *in situ* conservation of crop genetic resources. In: Brush SB (ed) Genes in the field: on-farm conservation of crop diversity. IPGRI, IDRC, Lewis Publishers

Butrón A, Tarrío R, Revilla P, Ordás A, Malvar R (2005) Molecular changes in the maize composite EPS12 during selection for resistance to pink stem borer. Theor Appl Genet 110: 1044-1051. doi:10.1007/s00122-005-1923-x

Ceccarelli S, Grando S (2007) Decentralized-participatory plant breeding: an example of demand driven research. Euphytica 155: 349–360. doi: 10.1007/s10681-006-9336-8

Ceccarelli S, Grando S (2007) Decentralized-participatory plant breeding: an example of demand driven research. Euphytica 155: 349–360. doi: 10.1007/s10681-006-9336-8

Cleveland DA, Soleri D, Smith SE (1999) Farmer plant breeding from a biological perspective: implications for collaborative plant breeding. CIMMYT Economics Working Paper No.10. CIMMYT, Mexico, DF Cooper D, Hodgkin T, Spillane C (2001) Broadening the genetic base of crops: an overview. In: Cooper D, Hodgkin T, Spillane C (eds) Broadening the genetic base of crop production. FAO, IPGRI, CABI

Emerson RA (1912) Inheritance of certain "abnormalities" in maize. Am Breed Assoc Rept 8: 385–399

Ferrão JEM (1992) A aventura das plantas e os descobrimentos portugueses. Programa Nacional de Edições Comemorativas dos Descobrimentos Portugueses, Portugal

Galinat WC (1980) Indeterminated *versus* determinated years. Maize Genet Coop Newsl 54: 121

Hallauer AR (1992) Recurrent selection in Maize. In: Janick J. (ed) Plant breeding reviews, vol 9. John Wiley & Sons, Inc, pp 115–177

Hallauer AR (1994) Corn genetics and breeding. Encyclopaedia Agric. Sci. 1: 455-467.

Hallauer AR, JB Miranda FO (1988) Quantitative genetics in maize breeding, 2nd edn. Iowa State Univ Press, Ames

Hallauer AR, Ross AJ, Lee M (2004) Long-term divergent selection for ear length in maize. In Janick J (ed) Plant breeding reviews, vol 24(2). John Wiley & Sons, Inc, pp 153–168

Hallauer AR, Russell WA, White PR (2000) Registration of BS21(R)C6 and BS22(R)C6 maize germplasm. Crop Sci 40:1517

Hallauer AR, Sears JH (1969) Mass Selection for Yield in Two Varieties of Maize1. Crop Sci 9: 47-50. doi: 10.2135/cropsci1969.0011183X000900010016x

IBPGR (1991) Descriptors for maize, Mexico City. International Board for Plant Genetic Resources, Rome

IPGRI (2000) Descritores para o milho, Mexico City. International Plant Genetic Resources Institute, Rome

Lopez-Reynoso JJ, Hallauer AR (1998) Twenty-seven cycles of divergent mass selection for ear length in maize. Crop Sci 38: 1099–1107. doi:10.2135/cropsci1998.0011183X003800040035x

Machado AT, Fernandes MS (2001) Participatory maize breeding for low nitrogen tolerance. Euphytica 122: 567– 573. doi: 10.1023/A:1017543426136

Moreira PM (2006) Participatory maize breeding in Portugal. A case study. Acta Agronomica Hungarica 54: 431–439. doi: http://dx.doi.org/10.1556/AAgr.54.2006.4.6

Moreira PM, Pêgo S (2003) Pre-breeding evaluation of maize germplasm. The case of a Portuguese open-pollinated variety. In: Abstracts of the Arnel R. Hallauer International Symposium on Plant Breeding. Mexico City, Mexico 17–22 August 2003

Moreira PM, Santos JP, Simões P, Santos JP, Vaz Patto MC, Carvalho V, Pêgo S (2005a) Pré-avaliação de populações de milhos regionais da região centro. A utilização do método «HUNTERS». II Colóquio de Melhoramento de Plantas e Conservação de Recursos Genéticos. Santarém 18 de Novembro. Escola Superior Agrária de Santarém.

Moreira PM, Santos JP, Simões P, Santos JP, Vaz Patto MC, Carvalho V, Pêgo S (2005b) Pré-avaliação de Populações de Milhos Regionais da Região Centro. Parâmetros Biométricos e Fitossanitários. VII Encontro Nacional de Protecção Integrada. 6 a 7 de Dezembro. Coimbra.

Pêgo S (1982) Genetic potential of Portuguese maize with abnormal ear shape, Ph.D. Thesis, Iowa State Univ.

Pêgo S (1996) Maize genetic resources in Portugal. In: Lipman E, Ellis RH, Gass T (eds), Maize genetic resources in Europe. Report of a workshop, Rome, Italy, IPGRI, pp 52-54.

Pêgo SE, Antunes MP (1997) Resistance or tolerance? Philosophy, may be the answer. In: Proceedings of the XIX – Conference of the International Working Group on Ostrinia. Guimarães Portugal 30th August–5th September 1997

Pêgo SE, Hallauer AR (1984) Portuguese maize germplasm with abnormal ear shape. Maydica 29: 39–53

Powell D (2002) The milling landscape of Northwest Portugal, Master's Thesis, California State University Northridge

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

Salazar R (2001) Empowering farmers and broadening the genetic base: agricultural research and resource management. In: Cooper D, Hodgkin T, Spillane C (eds) Broadening the genetic base of crop production. FAO, IPGRI, CABI

Sperling L, Ashby J, Weltzien E, Smith M., McGuire S (2001) Base-broadening for client-oriented impact: insights drawn from participatory plant breeding field experience. In: Cooper D, Hodgkin T, Spillane C (eds) Broadening the genetic base of crop production. FAO, IPGRI, CABI

Sthapit B, Friis-Hansen E (2000) Concepts and rationale of participatory approaches to conservation and use of plant genetic resources. In: Friis-Hansen E, Sthapit B (eds) Participatory approaches to the conservation and use of plant genetic resources. International Plant Genetic Resources Institute, Rome Italy

Sthapit B, Sajise P, Jarvis D (2005) Community based on-farm conservation of agricultural biodiversity: Good practices and lessons learned from Nepal and Vietnam. Second Colloquium on Plant Breeding and Plant Genetic Resources Conservation organised by Portuguese Association of Horticulture (APH) in Santarém City, Portugal, 18 November 2005.

Taba S, Díaz Rivas M, Rodríguez, Vicarte V, Norgaard J (2003) The CIMMYT Maize collection: Evaluation of accessions and Preliminary Breeder Core Subsets. In: Abstracts of the Arnel R. Hallauer International Symposium on Plant Breeding. Mexico City, Mexico 17–22 August 2003

Troyer AF (2000) Origins of modern corn hybrids. Proc Ann Corn Sorghum Res Conf 55: 27–42

Vaz Patto MC, Moreira PM, Almeida N, Šatović Z, Pêgo S. 2008. Genetic diversity evolution through participatory maize breeding in Portugal. Euphytica 161:283-291. doi: 10.1007/s10681-007-9481-8

Vaz Patto MC, Moreira PM, Carvalho V, Pêgo S (2007) Collecting maize (*Zea mays* L. convar. *mays*) with potential technological ability for bread making in Portugal. Genet Res Crop Evol 54:1555‑1563. doi: 10.1007/s10722-006-9168-3

Witcombe JR (2001) The impact of decentralized and participatory plant breeding on the genetic base of crops. In: Cooper D, Hodgkin T, Spillane C (eds) Broadening the genetic base of crop production. FAO, IPGRI, CABI

CHAPTER V.

The farmers' / breeders' selection dilemma revisited by long term participatory 'Pigarro' maize breeding analysis

The work presented in this chapter was mostly performed by Pedro Mendes Moreira (see acknowledgments section) and partially included in the following publication: Mendes-Moreira P, Šatović Z, Mendes-Moreira J, Santos, JP, Santos JPN, Pêgo SE, Hallauer AR, Vaz Patto MC. The farmers' / breeders' selection dilemma revisited by long term participatory 'Pigarro' maize breeding analysis. Paper under Submission

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
/UNL
Knowledge Creation

V.1. Abstract

This study is the continuation of Chapter IV and it refers to the analysis of a long term contribute of traditional genetic resources improvement on-farm using farmers and breeders selection methodologies for a more sustainable agriculture. For this analysis we have used a total of 159000 data points measured at plot and ear level, 111000 of which have been collected as new data for this ms. Additionally the molecular analysis of breeder selection was also included.

With this study we show for the first time that during both selection approaches of this participatory plant breeding project, genetic diversity changed to allow the maize population to phenotypically respond to selection, but was not reduced even with the most intensive breeder's selection. This diversity maintenance is providing to this already improved population the necessary resilience to further adapt to changing environments and alternative management practices. We conclude also that methods choice depends on the participatory breeding program main objectives: Phenotypic recurrent selection is easier and cheaper to adopt by farmers on OPV improvement, whereas breeder's selection results in a more uniform crop, being more adapted to hybrid development programs.

V.2. Introduction

Since its introduction, more than five centuries ago, maize has transformed the Portuguese agricultural panorama with many locally adapted maize landraces (Moreira 2006). In the 1960's, the Portuguese maize breeders, conscious of the threat to this unique national maize germplasm caused by diffusion of hybrids, started a regional collection of maize germplasm. More than 3000 accessions were collected and stored at the national plant germplasm bank, BPGV (Pêgo 1996), providing the basis for much of the national maize breeding achievements. Some of these achievements were attained through the participatory maize breeding "VASO" project (Sousa Valley project, initiated in 1984), implemented to answer to small farmers' concerns, such as how to increase yield without losing quality for bread production or ability for production in sustainable polycropping systems. The 'Pigarro' landrace was one of the landraces improved within this project, showing a strong ear fasciation expression. Fasciation can influence yield, being quite common among Portuguese traditional maize landraces (Vaz Patto et al. 2007).

Participatory Plant Breeding (PPB) has provided solutions for climate changes (Ceccarelli et al. 2013), diversity conservation (Maxted et al. 2002), organic and low input agriculture (Serpolay-Besson et al. 2014), and polycrop and agroecologic systems (Machado et al. 2011). PPB encourages interaction among plant breeders, other researchers and farmers, with the objective of developing cropping systems that

better meet local needs (Cleveland 2000). Several selection approaches, with different levels of farmers' involvement, can be found in PPB projects. In the case of 'Pigarro' participatory improvement, two selection approaches were applied: a farmer's phenotypic recurrent selection and a breeder's recurrent S2 lines selection.

In the previous study we compared the evolution of 'Pigarro' morphological response to farmer's and breeder's selection approaches, assessing just a few cycles of selection evaluated during two years of field trials (Mendes Moreira et al. 2008). At the molecular level, response to selection was assessed only at farmer's selection cycles (Vaz Patto et al. 2008). A more detailed comparative evaluation of the responses to selection at the phenotypic and genotypic levels is lacking to define the most effective and appropriate approach for a sustainable PPB. To fulfill this gap we conducted two more years of comparative farmer's *versus* breeder's selection cycles field trials. Molecular screening was also applied to the breeder selection cycles allowing a detailed comparison of both selection methods at agronomic, phenotypic, and molecular level.

Objectives of this study were to determine: 1) if 'Pigarro' initial population (from 1984) changed significantly, at phenotypic and molecular levels, during this long-term PPB; 2) if the two selection methods led to the same breeding outputs; 3) if any of the two selection methods significantly changed genetic diversity; and 4)

which of the two selection methods is the most useful for supporting PPB in sustainable farming systems.

V.3. Results

At the phenotypic level, although a few traits have evolved in the same direction in both selection methods, farmer's selection was more effective in increasing fasciation related traits and cob weight, with an overall significant contribution for yield increase (Table V.1). In comparison, breeder's selection was more effective in achieving crop uniformity, plant and ear height reduction, and greater resistance to stalk lodging (Table V.1). In our study, we detected only an increase of yield as a result of farmer's selection. An ear fasciation increase by farmers' selection was also confirmed and this, contrasted with the breeder's selection output (0.21% and -0.39% for yield selection gain, respectively for farmer's and breeders' cycles) (Table V.1). In addition, during breeders' selection, and contrary to the farmers' selection outputs, kernels became heavier. Finally ears became heavier for both farmers' (between C0 84 versus FSC20 04) and breeders' (between BSC2-94 versus BSC3-98) selection, especially due to cob weight increase (R^2 = 0.81 and gain cycle/year = 1.48% in farmers', R^2 = 0.87 and gain cycle/year = 2.17% in breeders' for cob weight evolution according selection) (Table V.2, Table V.3).

Table V.1Linear regression and analysis of variation of breeding methodologies applied to 'Pigarro' since 1984, based on field trials agronomical evaluation. Estimates of linear regressiofin coefficient (b), standard errors, initial cycle prediction (Ĉ0), coefficients of determination (R^2) and % of gain per year (%Gain/Y) for Farmer's selection (20 cycles) and for Breeder's selection (3 cycles). ANOVA for Farmer's (FS) and Breeder's (BS) cycles of selection, Env-environment; Y-years.

		Farmer's Selection																Breeder's Selection															
Traits	Data/plot	b				Ĉ0	R^2	%Gain/Y	FS	Y	Env	rep(Env)	FS * Y	FS * Env	Y * Env	FS * Y * Env	b				Ĉ0	R^2	%Gain/Y	BS	Y	Env	rep(Env)	BS * Y	BS * Env	Y * Env	BS * Y * Env			
50Fi	1	0.138	±	0.037	*	63.35	0.74	0.22	***	***	***	*			***		0.097	±	0.038		63.31	0.76	0.15	**	***	***	***	***	***	***	***			
50Ff	1	0.160	±	0.040	*	68.78	0.76	0.23	***	***	***	**		*	***	*	0.151	±	0.037		68.53	0.89	0.22	**	***	***				***				
50Mi	1	0.123	±	0.030	**	60.43	0.77	0.20	***	***	***	*			***		0.064	±	0.022		60.66	0.81	0.11		***	**				***				
50Mf	1	0.151	±	0.032	**	65.90	0.82	0.23	***	***	***				***		0.084	±	0.032		66.15	0.78	0.13	**	***	***				***				
OI	1	0.002	±	0.003		0.43	0.04	0.37		***					***		-0.003	±	0.002		0.49	0.52	-0.59							*				
MO	1	0.080	±	0.021	*	27.71	0.74	0.29	*	***	*	*			***		-0.024	±	0.046		27.25	0.12	-0.09		***	**				***				
CWEW	1	0.003	±	0.000	**	0.22	0.88	1.20	***	***	***		*		***		0.003	±	0.001		0.21	0.89	1.36	***	*	***				***				
Yld	1	0.014	±	0.017		6.91	0.12	0.21	*		***				***		-0.027	±	0.007		6.87	0.87	-0.39		*	***				**				
U	1	-0.001	±	0.006		2.80	0.01	-0.04		***	*						-0.032	±	0.006	*	2.89	0.93	-1.10	*						*				
N	1	0.010	±	0.006		5.06	0.36	0.20	**	***			***			*	0.018	±	0.011		5.02	0.58	0.36		*									
T	1	0.011	±	0.004	*	6.39	0.64	0.18		***	**						-0.003	±	0.019		6.50	0.02	-0.05	*	***	*								
E	1	-0.005	±	0.003		5.27	0.31	-0.10		***	**				*		-0.015	±	0.013		5.19	0.39	-0.29	*	***	***				***				
R	1	0.000	±	0.000		0.02	0.04	0.86		***					***		0.000	±	0.000		0.02	0.28	1.37		***		*			***				
S	1	0.001	±	0.001		0.06	0.21	0.98		***	**				***		-0.001	±	0.001		0.07	0.44	-1.26		***	**		*	*	***				
H	20	0.514	±	0.309		231.45	0.36	0.22	***	***	***	*			***		-0.600	±	0.516		230.03	0.40	-0.26	*	***	***	*			***				
H1E	20	0.371	±	0.234		138.06	0.33	0.27	***	***	***	**			***		-0.527	±	0.474		134.88	0.38	-0.39	***	**	***				***				
L	20	-0.052	±	0.016	*	17.37	0.66	-0.30	***	***	***				***		0.057	±	0.044		17.59	0.46	0.33	***	***	***	*			***				
ED1	20	0.041	±	0.007	**	5.67	0.87	0.73	***	***	**	**			***		0.007	±	0.013		5.54	0.14	0.13	***	***	***	*			**				
ED3	20	0.054	±	0.010	**	4.62	0.87	1.17	***	***	*	***					-0.005	±	0.011		4.51	0.09	-0.12	*	**	**		*						
ED2	20	0.031	±	0.005	**	5.29	0.90	0.59	***	***	***	*			***		0.006	±	0.009		5.20	0.19	0.12	***	***	***	*	*		**				
ED4	20	0.030	±	0.004	***	4.23	0.92	0.70	***	***		**					0.002	±	0.004		4.19	0.08	0.04		***	*		*						
R1	20	0.242	±	0.049	**	17.53	0.83	1.38	***	***	**				*		-0.042	±	0.081		16.78	0.12	-0.25	***	***	*	**							
R2	20	0.256	±	0.048	**	16.61	0.85	1.54	***	***	*	*					-0.090	±	0.079		16.15	0.39	-0.56	***	**	*								

		Farmer's Selection															Breeder's Selection														
Traits	Data/plot	b				Ĉ0	R^2	%Gain/Y	FS	Y	Env	rep(Env)	FS * Y	FS * Env	Y * Env	FS * Y * Env	b				Ĉ0	R^2	%Gain/Y	BS	Y	Env	rep(Env)	BS * Y	BS * Env	Y * Env	BS * Y * Env
Fa	20	0.064	±	0.015	**	1.94	0.78	3.31	***		***						-0.025	±	0.022		1.80	0.39	-1.36	***	*	*				*	
DI	20	-0.005	±	0.002	*	1.25	0.58	-0.43	***	**							-0.012	±	0.000	**	1.29	1.00	-0.96	***	***		*			**	
CV	20	0.027	±	0.008	*	1.86	0.71	1.47	***	***				*	*		-0.008	±	0.003		1.88	0.78	-0.41		*				*		
ECWEW	20	0.001	±	0.000	**	0.15	0.79	0.75	***	***		**	**	*	***		0.003	±	0.000	***	0.15	1.00	1.86	***	***	**		**		***	
EW	20	1.193	±	0.386	*	190.14	0.66	0.63	***	***	***	*			***		0.285	±	0.855		183.24	0.05	0.16	**	***	***	*			***	
KW	20	0.768	±	0.306		161.42	0.56	0.48	***	***	***	*			***		-0.297	±	0.698		156.38	0.08	-0.19	**	***	***	*			***	
CW	20	0.425	±	0.093	**	28.72	0.81	1.48	***	***	***	**			***		0.582	±	0.157		26.86	0.87	2.17	***	***	***	*	**		***	
Emo	20	0.005	±	0.008		15.95	0.07	0.03		***	***				***		-0.024	±	0.019		16.02	0.45	-0.15	*	***	***				***	
KD	20	0.000	±	0.000		1.01	0.05	-0.02		***	**	*			**		-0.002	±	0.001		1.00	0.57	-0.22	***	***	***				**	
SW	20	-1.606	±	0.354	**	347.78	0.80	-0.46	***	***	**	*			***		0.490	±	0.512		352.55	0.31	0.14		***	**	**	**		***	
KNº	20	5.039	±	1.369	*	465.29	0.73	1.08	***	***	***	*			***		-1.559	±	2.402		446.17	0.17	-0.35	***	***	***	*			***	
KR	20	-0.048	±	0.035		29.34	0.27	-0.16	***	***	***						-0.005	±	0.043		29.36	0.01	-0.02		***	***				***	
CD1	20	0.043	±	0.007	**	4.14	0.89	1.03	***	***	**	*			***		0.015	±	0.010		4.02	0.50	0.37	***	***	**				**	
CD3	20	0.054	±	0.009	**	3.18	0.87	1.68	***	***	*	**				*	0.000	±	0.014		3.07	0.00	-0.01	***	***	**					
CD2	20	0.030	±	0.004	***	3.63	0.91	0.82	***	***	**				***		0.015	±	0.004		3.57	0.87	0.42	***	***	**				**	
CD4	20	0.025	±	0.003	***	2.71	0.94	0.91	***	***		*					0.007	±	0.004		2.71	0.65	0.26	*	***	*					
M1	20	0.028	±	0.005	**	2.11	0.88	1.31	***	***	***				***		0.009	±	0.006		2.05	0.52	0.42	**	***	**	*			**	
M2	20	0.017	±	0.002	***	1.61	0.92	1.05	***	***	***				***		0.006	±	0.002		1.60	0.77	0.38		***	**	**			***	
Rq1	20	0.037	±	0.006	**	3.24	0.90	1.13	***	**	***				***		0.015	±	0.006		3.14	0.77	0.49	***						***	
Rq2	20	0.026	±	0.003	***	2.70	0.93	0.96	***	***	***				***		0.015	±	0.001	**	2.67	0.99	0.56	***	*	**				***	

* - Significant at 0.05 probability levels; ** - Significant at 0.01 probability levels; *** - Significant at 0.001 probability levels; 50Fi - Days-to-silk beginning, nº; 50Ff - Days-to-silk end, nº; 50Mi - Days-to-anthesis beginning, nº; 50Mf - Days-to-anthesis end, nº; OI - Overlap Index between beginning and end of anthesis and silking; MO – Moisture, %; CWEW – Cob and ear weight ratio at harvest; Yld - Yield, Mg ha^{-1}; U - Uniformity; N - aNgle; T - Tassel; E - Ear placement; R - Root lodging, %; S - Stalk lodging, %; H - Plant height, cm; H1E - Ear height, cm; L - Ear Length, cm; ED1 - Ear Diameter 1, cm; ED3 - Ear Diameter 3, cm; ED2 - Ear Diameter 2, cm; ED4 - Ear Diameter 4, cm; R1 - Kernel-row number 1, nº; R2 - Kernel-row number 2, nº; Fa – Ear fasciation; DI - determinate versus indeterminate ears; CV - Convulsion; ECWEW - Cob/Ear weight; EW - Ear weight, g; KW - Kernel weight, g; CW - Cob weight, g; Emo - Ear Moisture, %; KD - Kernel dept, cm; SW - Thousand kernel weight, g; KNº - Kernel number, nº; KR - Kernel per row, nº; CD1 - Cob diameter 1, cm; CD3 - Cob diameter 3, cm; CD2 - Cob diameter 2, cm; CD4 - Cob diameter 4, cm; M1 - Medulla 1, cm; M2 - Medulla 2, cm; Rq1 - Rachis 1, cm; Rq2 - Rachis 2, cm (trait detailed information in Mendes-Moreira et al., 2008).

Molecular results confirm that both breeding approaches seem to have achieved phenotypic modifications though preserving genetic diversity. The lack of significant differences among farmer's and/or breeder's selection cycles in any of the diversity parameters analyzed (N_{ar}, H_O, H_E, f) (Table V.2, Table V.3) indicates no effective loss of genetic diversity occurring during the two selection methods. Of the 81 different originally detected alleles, using 15 SSR markers, 61 alleles were maintained in FSC20-04 and 59 alleles were maintained in BSC3-98, reinforcing the idea of that genetic variability was maintained (Figure V.1, Table V.2, Table V.3). In addition, the number of common/shared alleles among selection cycles was 75.31% and 72.84% for farmer's and breeder's selection, respectively.

Table V.2 Genetic variability estimates for the initial population (C0-84), three breeder's selection cycles (BSC1-89, BSC2-94, BSC3-98) and two farmer's selection cycles (FSC9-93, FSC20-04).

Selection cycle	n	N_a	N_{ar}	N_{pa}	H_O	H_E	f
C0-84	30	5.400	3.718	8	0.483	0.584	0.176
BSC1-89	30	4.800	3.348	2	0.442	0.547	0.195
BSC2-94	30	4.933	3.522	3	0.469	0.592	0.212
BSC3-98	30	4.800	3.760	2	0.552	0.652	0.156
FSC9-93	29	4.667	3.409	1	0.570	0.588	0.032
FSC20-04	30	4.733	3.503	3	0.509	0.597	0.153
Average		4.889	3.543		0.504	0.593	0.154
P(KW)[#]			0.729		0.219	0.654	0.682
P(BSC vs. FSC)[†]			0.598		0.317	0.917	0.065

[#]Probability of Kruskal-Wallis test among all selection cycles

[†]P-value of the permutation tests for difference between selecting methods (BSC vs. FSC)

n: number of individuals, N_a: average number of alleles, N_{ar}: allelic richness, N_{pa}: number of private alleles, H_O: observed heterozygosity, H_E: gene diversity or expected heterozygosity, f: inbreeding coefficient.

Table V.3 AMOVA for partitioning of SSR variation between selection methods (Breeder's vs. Farmer's), among cycles within selection methods and within selection cycles

Source of variation	% Total variance		Φ-statistics	PΦ)
	Between	Within	Φ	
Breeder vs. Farmer selection methods	2.43		Φ_{CT} = 0.024	< 0.001
Among cycles within selection methods		5.16	Φ_{SC} = 0.053	< 0.001
Within cycles		92.40	Φ_{ST} = 0.076	< 0.001
All cycles	6.40	93.60	0.064	< 0.001
Breeders' cycles*	6.77	93.23	0.068	< 0.001
C0-84 vs BSC1-89	8.75	91.25	0.087	< 0.001
BSC1-89 vs BSC2-94	6.04	93.96	0.060	< 0.001
BSC2-94vs BSC3-98	4.53	95.47	0.045	< 0.001
C0-84 vs BSC3-98	5.52	94.48	0.055	< 0.001
Farmers' cycles*	3.24	96.76	0.032	< 0.001
C0-84 vs FSC9-93	2.62	97.38	0.026	< 0.001
FSC9-93 vs FSC20-04	4.07	95.93	0.041	< 0.001
C0-84 vs FSC20-04	3.03	96.97	0.030	< 0.001

#Probability of Kruskal-Wallis test among all selection cycles; †P-value of the permutation tests for difference between selecting methods (Breeder's vs. Farmer's);
*Comparisons of both Breeder's and Farmers' cycles include the initial population (C0-84); P(ϕ) - ϕ-statistics probability level after 10,000 permutations.

AMOVA analysis among selection cycles also indicated a greater proportion of genetic diversity maintained within each selection cycle; 94.48% and 96.97% of the variation was attributable to within-selection cycles diversity for breeders' selection and farmers' selection, respectively (Table V.2, Table V.3). In addition, this analysis also showed that the percentage of total variance among cycles within selection methods per se (5.16%), was two times greater than between selection methods (2.43%).

Factorial correspondence analysis indicated, along its first axis, two different genetic directions for the two selection methods (Figure V.2). The first farmer selection cycle analyzed, FSC9-93, was however closer to the breeder's selection. This corresponded with a more

stratified mass selection applied since 1986 until 1999. More recent farmer's selections were much more differentiated from the breeders and more differentiated among them. Along the FCA second axis, a major distance between the final farmers' cycle analyzed, FSC20-04, and the original population, was observed. Breeder's selection gave rise to much more uniform populations than farmer's selection (Table V.2, Table V.3, Figure V.2).

Allele frequency distributions have changed significantly between selection cycles for a few of the *loci* under evaluation (data not shown). The number of private alleles, however, varied among selection cycles, being, as expected, the highest in the original population (Figure V.2). We observed that locus umc1907 significantly deviated from Hardy-Weinberg equilibrium ($P< 0.05$) in all selection cycles (farmer's and breeder's), umc1823 only for the breeder's selection and umc1229 only for the farmer's selection cycles.

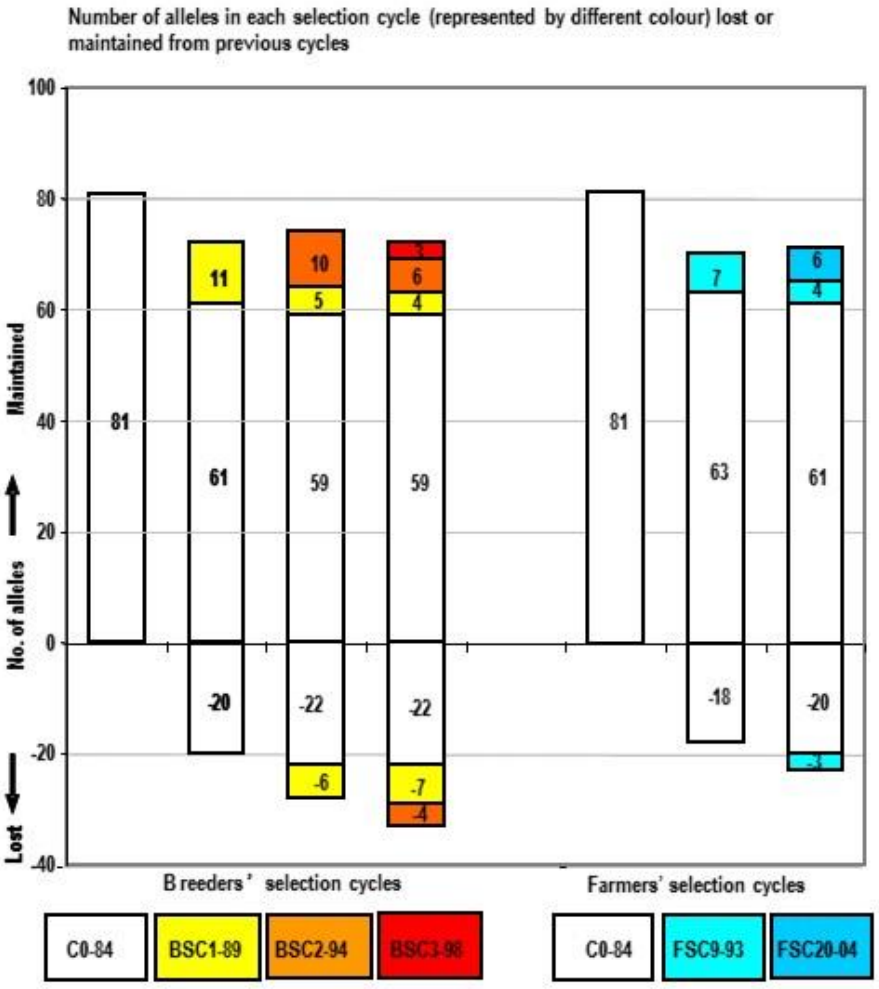

Figure V.1 Number of alleles in each selection cycle lost or maintained from previous cycles detected using 15 SSRs markers. Negative numbers refer to alleles lost comparing with previous analysed cycle. Positive numbers refer to new alleles or alleles maintained comparing with previous analysed cycle.

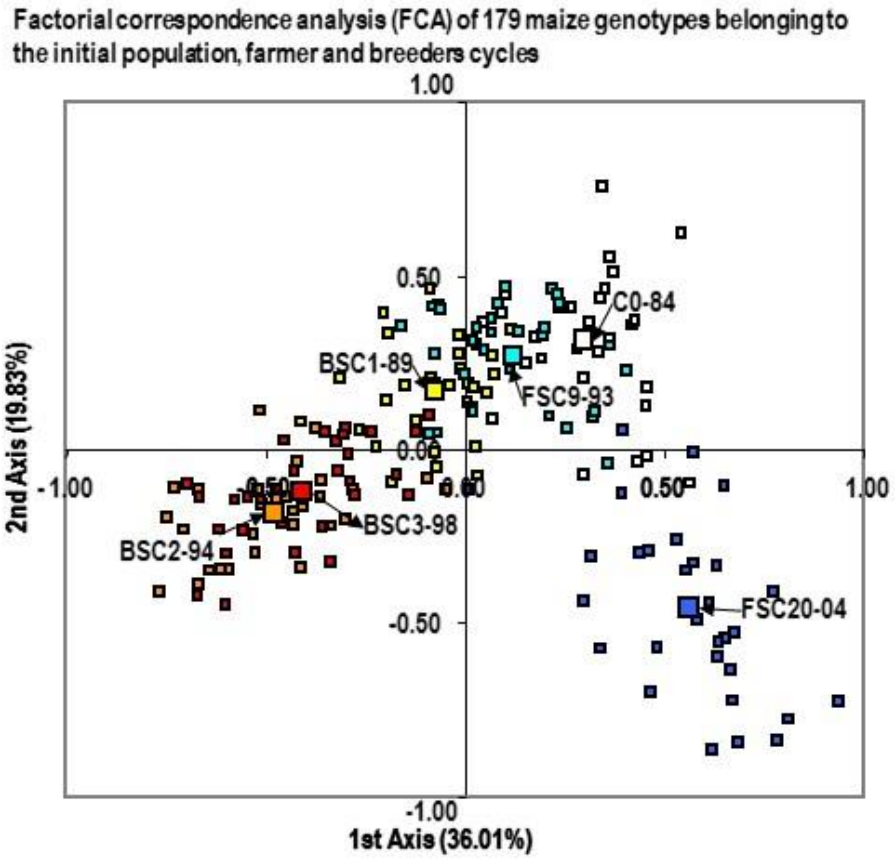

Figure V.2 Factorial correspondence analysis (FCA) of 179 maize genotypes belonging to the initial population (C0-84), three breeder's selection cycles (BSC1-89, BSC2-94, BSC3-98) and two farmer's selection cycles (FSC9-93, FSC20-04). Each individual genotype is indicated by a small symbol, while the population barycentres are represented by larger symbols.

V.4. Discussion

The maize 'Pigarro' population was under selection since 1985, within the PPB VASO project, using a farmer's and a breeder's approach. To identify the most useful selection approach to support participatory maize breeding in sustainable farming systems, we

compared ´Pigarro´ molecular diversity evolution and agronomic selection response between the two applied selection approaches. We confirmed that during both selection approaches, genetic diversity changed, to allow the 'Pigarro' population to phenotypically respond to selection. Nevertheless, genetic diversity was not reduced even with the more intensive breeder's selection, suggesting further response to selection can be expected.

The evaluation of both selection methods suggested that both selection approaches were effective for achieving the main breeding objectives. As an example, crop uniformity was significantly improved by breeder selection (R^2 = 0.93 and gain per year 1.10%), but not by farmer's selection. Uniformity is important for hybrid development and to comply with seed commercialization requirements. In our study we only detected yield increase during farmer's selection. Increased ear fasciation might be partially responsible for this observed yield improvement. The ear fasciation increase by farmer selection was reported previously by Mendes Moreira et al. (2008, 2015). Ear fasciation is a particularly important trait for farmers during their seed selection, where they balance the choice of fasciated ears with other ear types to maintain a certain level of diversity, towards a long term gain in ear diameters, kernel row numbers, medulla and rachis dimensions (Vaz Patto et al., 2007). This positive selection of fasciation by farmers, contrast with breeder's selection suggesting an important role of fasciation for yield improvement. In case of breeder's selection, yield improvement

strategy can be probably associated with adaptation to increased plant densities, considering that during breeder's selection it was observed a reduction of plant height and yield.

During farmers' selection an increasing level of kernel convulsion and the number of kernels per ear was associated with a decrease of thousand kernel weight, indicating a reduction of kernel size. In parallel, during this selection, ear length decreased significantly, and kernel row number as well as ear diameter increased, in agreement with Emerson and East (1913) and Hallauer et al. (2010) for the long-term divergent selection of ear length in maize. Nevertheless, contrary to Hallauer et al. (2010), yield slightly increased even though ear length was reduced. During breeders' selection, contrary to the farmer's selection outputs, kernels became heavier, indicating a tendency for bigger kernels, considering that the kernel type did not change.

Ear weight increase maybe highly demanding for stalk lodging resistance and root anchorage, being because of this potentially associated with lower values of stalk and root lodging respectively. However, this association was only observed at the farmer's selection, with a high correlation between root or stalk lodging with cob weight (r=0.529; 0.234) and with cob and ear weight ratio (r=0.573; 0.266). The observed higher correlations between cob/ear weight ratio at harvest and per ear, with medulla and rachis 1 and 2 (data not shown), suggested a higher lignification of the rachis, which may be important for ear architecture regarding kernel support.

Mendes Moreira et al. (2008), stated that differences in yield response between both selection methods could be related to a reduction in diversity along breeder's selection. Concerns have also been expressed that genetic diversity may be reduced by natural and artificial (human) selection (Vaz Patto et al., 2008).

Genetic differentiation for breeder's selection cycles decreased progressively with cycle increase, while during farmer's selection, genetic differentiation changed more erratically, being higher between FSC9-93 and FSC20-04 (4.07%) than between C0-84 and FSC9-93 (2.62%) (Table V.2, Table V.3). This difference can be associated with changes reported on the farmer's selection objective since 1993 (beginning of "Sousa Valley Best Ear" competition) towards increased ear sizes.

Changes observed in allelic frequency distribution and number of private alleles suggested that genetic diversity has not been reduced from 'Pigarro' population in 1984 to those improved by farmer's or breeder's selection, but the genetic diversity maintained was not exactly the same. These molecular changes, and depending on the selection approach, also had a phenotypic expression according to the previously discussed phenotypic data evolution. Considering that the mutation rate in maize is generally very low (Kahler et al. 1986) and the seed maintenance procedure used during this PPB selection was by isolation plantings and a farmer's or breeder's selection pressure of 1-5% or 15-20% respectively (Mendes Moreira et al. 2008), it is expected that assortative mating and/or selection were

the most likely reasons for explaining deviations from the Hardy-Weinberg equilibrium. In the case of breeder's selection, possible inbreeding effects could have also contributed to the observed deviations from the Hardy-Weinberg equilibrium.

The majority of the screened SSR *loci* represented non-coding DNA regions (9 of the 15 SSR markers used were genomic SSRs) apparently not subject to strong selection pressures (Heath et al. 1993). However, they could be linked to selected *loci* and therefore subjected to selection by genetic hitchhiking (Pinto et al. 2003). This suggests that directional selection observed on these SSR markers might indicate *loci* controlling the selected trait or traits linked to these markers (Butrón et al. 2005). Indeed, after accounting for multiple comparisons, several SSR *loci* were out of Hardy-Weinberg equilibrium in a few of the selection cycles. Due to space constraints we will only refer to the ones consistently selected across improvement cycles. These were all genomic SSRs and all with an excess of homozygotes. In particular, umc1907, significantly out of Hardy-Weinberg equilibrium ($P < 0.05$) in all selection cycles (farmer's and breeder's), is located at maize genome bin 3.05, where several genes have been identified that might be associated with or indicating *loci* controlling traits consistently changing in both selection approaches. The candidate gene *terminal ear1* (*te1*) and several QTLs controlling days to pollen 2, 7, 12 (qdpoll2, 7, 12) were detected in this region. Days to pollen or anthesis were found to be inversely correlated with the average of determinate *versus*

indeterminate ears ears (<-0.72 for farmer's selection and <-0.89 for breeder's selection) and are mainly associated with cycle duration. Indeed in both selection methods plant cycles tended to increase and ears became more indeterminate. In addition, the QTL ear diameter7 (qeard7) that can be associated with the ear diameters phenotype genetic control, was also located in this region. The majority of the detected correlations between ear and cob diameters were higher than 90%, although Hallauer et al. (2010) reported 67% and Mendes-Moreira et al. (2015.) reported 80.7%. These high correlations may be associated with *loci* controlling the cob diameters increase with both selection methods and cycles.

Locus umc1823, significantly deviated from Hardy-Weinberg equilibrium only for the breeder's selection cycles, is located at bin 2.02, where several genes potentially associated with traits consistently changing with the breeder's selection have been identified. This is the case of the QTLs for cob diameter 14 (qcobd14) and kernel row number 6, 26 (qkrow6, 26). Indeed, very high correlations among cob diameter3 and row number1, with ear diameter3, have been described in the present study and by others (>0.90, on this study or >0.80 by Mendes-Moreira et al. (2015)). These traits can be associated with *loci* controlling ear length increase, and the reduction of ear fasciation and kernel depth observed with the breeder's selection. In addition, fasciation was in this study correlated with cob diameter 3 (0.78). Mendes-Moreira et

al. (2015) had also indicated a correlation of 0.59 to 0.79 among the same traits at two different locations.

Finally, umc1229, significantly deviated from Hardy-Weinberg equilibrium only for the farmer's selection cycles, is located at bin 6.01, where several genes, potentially associated with phenotypic traits consistently changing along the farmer's selection, have been identified. With farmer selection, ears become shorter and wider, with a greater number of rows. Hence with more convulsion, higher fasciation and smaller kernels that increased in number. Among bin 6.01 potentially associated genes we may find the *defective kernel19*, *28* (*dek19, 28*), and *miniature seed3* (*mn3*) genes, associated with the observed decrease of thousand kernel weight. On this same region, the ear length 25 (qearl25) and days to pollen4 (qdpoll4) QTLs were also detected (Lawrence et al., 2005). The qearl25 might be associated with the observed ear length decrease, while the qdpoll4 might be associated with the observed ear and cob diameter increases, due to the high correlation detected among these traits and the beginning of anthesis, i.e., days to pollen (>0.85, in Mendes-Moreira et al. (2015)).

To identify the most useful selection approach to support participatory maize breeding in sustainable farming systems, we compared ´Pigarro´ maize OPV agronomic selection response and evolution of molecular diversity in two distinct selection methods; farmer's phenotypic recurrent selection and breeder's S2 lines recurrent selection.

In conclusion, we confirmed that during both selection approaches, genetic diversity changed, to allow the population to phenotypically respond to selection, but was not reduced even with the most intensive breeder's selection. Although there were no significant differences detected on the studied genetic diversity parameters along selection cycles, during both selection methods, an increase in plant maturing and in the ears indeterminacy was observed. Also in both selection methods, cobs have become wider and heavier. The last cycle of both selection methods maintained the ability for polycropping systems and quality for bread production according to Vaz Patto et al. (2009, 2013). Nevertheless, particular phenotypic traits evolved in opposite directions between the two selection methods. With breeder's selection ears became longer and less fasciated with an overall increase of crop uniformity, whereas farmer's selected for shorter and wider ears, with increased levels of fasciation and smaller kernels. Our molecular diversity evolution analysis highlighted potential associations between particular neutral molecular markers and *loci* controlling some of the phenotypic traits under selection (*e.g.*, ear length, fasciation and related ear traits as ear diameter and kernel-row number). These associations need however to be better explored and validated by future linkage or association mapping approaches previous to their use for supporting trait selection in sustainable farming systems.

V.5. Material and Methods

V.5.1 Germplasm development

'Pigarro' is a FAO 300 maturity Open Pollinated Variety (OPV) with white flint kernels, high levels of root and stalk lodging and high kernel-row numbers (normally between 18 and 28, but 48 rows have already been observed). Its improvement, since 1985 under the VASO project breeding approach, focused on two main recurrent selection methodologies: farmer's selection and breeder's selection. Farmer's selection (FS) corresponded to a phenotypic recurrent selection using stratified mass selection, with two parental control in three sequential steps: 1) negative selection by detasseling before anthesis; 2) plant and ear selection, based on stalk quality and ear size; and 3) best ears selection at storage facilities. Breeder selection (BS) corresponded to a S2 lines recurrent selection, considering the additive component of genetic variance (i.e., 3/2 σ_a^2 *versus* σ_a^2, respectively, for S2 and S1 lines) (Hallauer et al., 2010), organized in a four season scheme. Both selection procedures are described in detailed by Mendes-Moreira et al. (2008).

Both methods emphasized selection for yield, pest and diseases reaction and indirectly quality for maize bread (Vaz Patto et al. 2009; 2013).

Seed from each selection cycle of 'Pigarro' VASO Project, either from farmer's or breeder's selection, were stored at 4ºC in NUMI (Maize Breeding Station, Braga, Portugal) cold storage facilities.

V.5.2 Phenotypic evaluation

To determine the effectiveness of both methods of selection, seed from both farmer's selection (six cycles: FSC4-88, FSC6-90, FSC9-93, FSC12 96, FSC15-99, FSC20-04) and breeder's selection cycles (three cycles: BSC1 89, BSC2 94, BSC3 98), and the initial 'Pigarro' population (C0-84), were included in comparative field trials. Field trials were established at three locations in Portugal (Coimbra 40°13'0.22"N, 8°26'47.69"W; Montemor 40°10'4.82"N, 8°41'14.84"W and Lousada 41°14'03.43"N, 8°18'13.11"W) during four years, from 2005 till 2008. However, extreme drought after sowing, in 2006 at Montemor, and late thinning, in 2008 at Lousada, restricted data collection at both sites. Coimbra and Montemor are in the river Mondego irrigation perimeter, a very high-yielding area where the average yield for maize hybrids is 14.5 $Mgha^{-1}$. Lousada is located in a traditional maize production region, with an average maize hybrid production of 8 $Mgha^{-1}$.

Sowing occurred in May, differing 15 days among locations, and harvests in October.

For each environment, a randomized complete block design, with three replications, was used. Each replication included two rows plot (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). Plots were overplanted by hand and thinned at the seven leaf stage (Ritchie et al. 1993), for a final stand of approximately 50,000 plants ha^{-1}. Plots

were mechanical and/or hand weeded as necessary and managed following common agricultural practices for maize in the region. All the plots were harvested by hand.

Phenotypic data were collected per plot or from a group of 20 plants and/or ears per plot, for 43 traits (Table V.1), as described by Mendes Moreira et al. (2008), with some minor changes. Uniformity score scales varied from one (minimum) to nine (maximum). In maize populations average values ranged from one (minimum) to a maximum of five, being this average values six to nine in inbred and hybrids. Cob/ear ratio at harvest was determined based on the measurement of five shelled ears.

V.5.3 Phenotypical data analysis

Data analysis was conducted separately for both selection methods. Analyses of variance were computed using IBM SPSS Statistic 22.0 for selection cycles, environments (locations), years, and the respective interactions with selection cycles. Replications were nested in environments.

Phenotypic data from 2005 and 2006 field trials and from 2005 ear traits were previously published (Mendes Moreira et al. 2008) and made available for this new comparative analysis.

Response to selection was evaluated for farmer and breeder selection using a linear regression model (Excel 2003), regressing observed populations means on cycle of selection (b=regression of

trait on cycle of selection and response expressed relative to the C0 population and on a year basis).

V.5.4 Molecular evaluations

For molecular comparison, the initial population (C0-84), the two farmer's (FSC9-93, FSC20-04) and the three breeder's selection cycles (BSC1-89, BSC2-94, and BSC3 98) were used. Molecular data for C0-84, FSC9-93, and FSC20-04 were published previously by Vaz Patto et al. (2008), named as SC1984, SC1993 and SC2004, and made available for this new analysis (Table V.2, Table V.3). For each analyzed cycle, 30 individuals were randomly selected from seed stocks.

DNA was isolated from a total of 90 individuals corresponding to the three breeder's selection cycles (using 2-week old seedling leaf samples), employing a modified CTAB procedure (Saghai-Maroof et al. 1984). These individuals were subsequently screened with the same 15 SSRs markers (umc1013, umc1823, umc1635, umc1907, umc1528, umc1524, umc1143, umc1229, umc1066, umc1483, umc1858, umc1279, umc1120, umc2067, umc2021) previously used in Vaz Patto et al. (2008) to allow comparisons. SSR marker technique was performed as in Vaz Patto et al. (2004). Fragment analysis was conducted using automated laser fluorescence (ALFexpress II) sequencer (Amersham Biosciences), as in Vaz Patto et al. (2008).

Amplification fragments size was determined in base pairs and visually scored at least twice independently for each entry, to ensure

data accuracy. Data from Vaz Patto et al. (2008) were added to this matrix (adding up to a total of 179 individuals) for the comparative analysis of all farmer's and breeder's selection cycles.

V.5.5 Molecular data analysis

Several genetic diversity parameters, such as Polymorphism Information Content (PIC), allele frequencies, average number of alleles (N_a), number of private alleles (N_{pa}), observed and expected heterozygosities (H_O, H_E), inbreeding coefficient (f) and allelic richness (N_{ar}), were calculated using the SSR data matrix, as in Vaz Patto et al. (2008).

The estimates of N_{ar}, H_O, H_E and f estimates in each selection cycle were compared using the Kruskal-Wallis test. Average values of Nar, H_O, H_E and f were tested for significant differences between breeder's and farmer's selection. Genotypic frequencies were tested for conformance to Hardy-Weinberg (HW) expectations, as well as to estimate the significance of genic differentiation between selection cycle pairs. Analysis of molecular variance (AMOVA) (Excoffier et al. 1992) was used to partition the total microsatellite diversity among and within groups defined by taking into account different selection methods and cycles. All these analysis were performed as in Vaz Patto et al. (2008). A factorial correspondence analysis (FCA) was carried out using Genetix 4.05 (Belkhir 2004), in order to graphically represent genetic relationships among individual genotypes.

V.6. Acknowledgements

Conception and design of the work: MCVP, ARH, SP

Acquisition of data: phenotypic (PMM, JPNS, JPPS) and molecular PMM

Analysis and interpretation of data: PMM, ZS, MCVP, JMM

Article drafting: PMM, MCVP

Revising it critically: ARH, SP, MCVP, JMM

Population development and breeding: SP, farmer FM.

V.7. References

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F. 2004. GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Page WEB de GENETIX. http://kimura.univ-montp2.fr/genetix/. Accessed 21 July 2014

Butrón A, Tarrío R, Revilla P, Ordás A, Malvar R (2005) Molecular changes in the maize composite EPS12 during selection for resistance to pink stem borer. Theor Appl Genet 110: 1044-1051. doi:10.1007/s00122-005-1923-x

Ceccarelli S, Galie A, Grando S (2013) Participatory Breeding for Climate Change-Related Traits. In: Chittaranjan K (ed) Genomics and Breeding for Climate-Resilient Crops. Springer, pp 331-376. doi:10.1007/978-3-642-37045-8_8

Cleveland DA, Daniela S, Smith SE. 2000. A biological framework for understanding farmers' plant breeding. Econ Bot 54: 377-394. doi: 10.1007/BF02864788

Emerson R, East E. 1913. The Inheritance of Quantitative Characters in Maize, Nebr. Agric. Expt. Sta. Bull 2.

Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to

Human Mitochondrial DNA Restriction Data. Genetics 131:479-491. http://www.genetics.org/content/131/2/479.abstract

Hallauer AR, Carena MJ, Miranda Filho JB (2010) Quantitative genetics in maize breeding. Springer, New York. doi:10.1007/978-1-4419-0766-0

Hallauer AR, Sears JH (1969) Mass Selection for Yield in Two Varieties of Maize1. Crop Sci 9: 47-50. doi: 10.2135/cropsci1969.0011183X000900010016x

Heath DD, Lwama GK., Devlin RH. 1993. PCR primed with VNTR core sequences yields species specific patterns and hypervariable probes. Nuc Acids Res 21:5782-5785. doi:10.1093/nar/21.24.578

Kahler A, Hallauer A, Gardner C. 1986. Allozyme polymorphisms within and among open-pollinated and adapted exotic populations of maize. Theor Appl Genet 72: 592-601. doi: 10.1007/BF00288996

Lawrence CJ, Seigfried TE, Brendel V. 2005. The Maize Genetics and Genomics Database. The Community Resource for Access to Diverse Maize Data. Plant Physiol 138: 55-58. doi:10.1104/pp.104.059196

Machado AT, Nass LL, Machado TTC (2011) Manejo sustentavel agrobiodiversidade nos biomas Cerrado e Caatinga. EMBRAPA.

Maxted N, Guarino L, Myer L, Chiwona E. 2002. Towards a methodology for on-farm conservation of plant genetic resources. Genet Resour Crop Ev 49: 31-46. doi: 10.1023/A: 1013896401710

Mendes Moreira PMR, Pêgo SE, Vaz Patto MC, Hallauer AR (2008) Comparison of selection methods on 'Pigarro', a Portuguese improved maize population with fasciation expression. Euphytica 163: 481-499. doi: 10.1007/s10681-008-9683-8

Mendes-Moreira P, Alves ML, Šatović Z, dos Santos JP, Santos JN, Souza JC, Pêgo SE, Hallauer AR, Vaz Patto MC. 2015. Genetic Architecture of Ear Fasciation in Maize (*Zea mays*) under QTL Scrutiny. PLoS ONE 10(4): e0124543. doi:10.1371/journal.pone.0124543

Mendes-Moreira PM, Mendes-Moreira J, Fernandes A, Andrade E, Hallauer AR, Pêgo SE, Vaz Patto M (2014) Is ear value an effective indicator for maize yield evaluation? Field Crop Res 161: 75-86. doi: 10.1016/j.fcr.2014.02.015

Moreira PM (2006) Participatory maize breeding in Portugal. A case study. Acta Agronomica Hungarica 54: 431–439. doi: http://dx.doi.org/10.1556/AAgr.54.2006.4.6

Pêgo SE (1996) Maize genetic resources in Portugal. In: Lipman E, Ellis RH, Gass T (eds), Maize genetic resources in Europe. Report of a workshop, Rome, Italy, IPGRI, pp 52-54.

Pinto LR, Vieira MLC, de Souza Jr CL, de Souza AP. 2003. Genetic-diversity assessed by microsatellites in tropical maize populations submitted to a high-intensity reciprocal recurrent selection. Euphytica 134: 277-286. DOI: 10.1023/B: EUPH.0000004946.15260.4a

Ritchie SW, Hanway JJ, Benson GO. 1993. How a Corn Plant Develop. Iowa State University of Science and Technology, Ames, Iowa.

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW. 1984. Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci 81: 8014-8018. doi: /10.1073/pnas.81.24.8014

Serpolay-Besson E, Giuliano S, Schermann N, Chable V (2014) Evaluation of Evolution and Diversity of Maize Open-Pollinated Varieties Cultivated under Contrasted Environmental and Farmers' Selection Pressures: A Phenotypical Approach. Open Journal of Genetics 4: 125-145. doi: 10.4236/ojgen.2014.42014

Vaz Patto MC, Alves ML, Almeida NF, Santos C, Mendes-Moreira P, Šatović Z, Brites C. 2009. Is the bread making technological ability of portuguese traditional maize landraces associated with their genetic diversity? Maydica 54: 297-311. http://www.maydica.org/articles/54_297.pdf

Vaz Patto MC, Mendes-Moreira PM, Alves ML, Mecha E, Brites C, do Rosário Bronze M, Pêgo S. 2013. Participatory Plant Quality Breeding: An Ancient Art Revisited by Knowledge Sharing. The Portuguese Experience. In: Andersen SB (ed) Plant Breeding from Laboratories to Fields, InTech. doi: 10.5772/52951

Vaz Patto MC, Moreira PM, Almeida N, Šatović Z, Pêgo S. 2008. Genetic diversity evolution through participatory maize breeding in Portugal. Euphytica 161:283-291. doi: 10.1007/s10681-007-9481-8

Vaz Patto MC, Moreira PM, Carvalho V, Pêgo S (2007) Collecting maize (*Zea mays* L. convar. *mays*) with potential technological ability for bread making in Portugal. Genet Res Crop Evol 54:1555‑1563. doi: 10.1007/s10722-006-9168-3

Vaz Patto MC, Šatović Z, Pêgo S, Fevereiro P (2004) Assessing the genetic diversity of Portuguese maize germplasm using microsatellite markers. Euphytica 137: 63-72. doi:10.1023/B:EUPH.0000040503.48448.97

CHAPTER VI.

Is ear value an effective indicator for maize yield evaluation?

The work presented in this chapter was mostly performed by Pedro Mendes Moreira (see acknowledgments section) and partially included in the following publication: Mendes Moreira, P.M., Mendes-Moreira, J., Fernandes, A., Andrade, E., Hallauer, A.R., Pêgo, S.E., Vaz Patto, M.C., 2014. Is ear value an effective indicator for maize yield evaluation? Field Crops Research 161, 75-86.

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
/UNL
Knowledge Creation

VI.1. Abstract

The present Chapter VI intend to answer the following questions 1)How can we test alternative interpretable regression methods (namely from multiple linear regression and multiple adaptive regression splines) to provide new ear value formulas that better estimates the yield potential using ear traits?; 2) Can we develop a new instance ranking method, allowing to select the best new ear value formula to be used on the ear competition?; 3) How can we identify a set of traits that will help farmers on selection towards better yield; and 4) How can we compare the ranking results obtained by the original EV formula and the newly one developed, using data from the "Sousa Valley Best Ear" competition.

To answer these questions we have used as a case study the Sousa Valley Best Ear Competition that takes place in Portugal (Paredes city) since 1992. An Ear Value formula, not directly associated with yield aspects, was developed, based on bibliographic correlations, for this competition. We trialed several cycles of Participatory Plant Breeding of some of the competition winning maize populations during three years in two to three locations. These trials allowed us to collect data on yield, field and ear traits. These data were analyzed based on Multiple Linear Regression (MLR) and Multiple Adaptive Regression Splines (MARS). Eleven methods for yield prediction were ranked based on a new ranking method (PR.NDCG measure). We have selected the most appropriate formula that included the original EV traits but with different coefficients and entitled as Adjusted EV

(EVA). Finally we have compared the ranks obtained with EV and EVA when data from the ear competition were used.

The interpretable derived models in this study were specific to the range of populations used in the competition of "Sousa Valley Best Ear". However, to some extent, such models can be calibrated for use with other maize populations. Furthermore they can be expanded to pre-breeding, on-farm conservation or to better understand breeding selection procedures evolution along time and from breeder to breeder. In addition it can be used as a tool in Participatory Plant Breeding (PPB) projects where quantitative information is collected by farmers in order to improve their own selection procedures.

VI.2. Introduction

VI.2.1 Context

Since maize (*Zea mays* L.) domestication from teossinte (*Z. mays* ssp. *parviglumis*) (Doebley 2004) 6000 to 10,000 years ago, farmers have selected according to multiple traits, such as kernel composition (*e.g.* sweet corn, starch type), palatability, speed of germination and stalk strength (Wilkes 2004). Selection of maize landraces by farmers is still a common practice in many countries in the world. Farmers' experience and perception has allowed translating maize physical traits into meaningful indicators of yield, insect resistance or simple esthetic value (Fitzgerald 1993). However, a precise description of their selection criteria is not always easy to obtain due to the use of

indirect measurements (Pressoir, Berthaud 2004; Badstue et al. 2007). Moreover, the selection traits are largely confined to ear characteristics, offering a limited scope for further variety improvement (Louette, Smale 2000).

In the beginning of the twentieth century, both the development of popular maize ear show exhibitions and the implementation of a scientific approach to maize inbred lines (Shull 1908, 1909), outlined the foundation of modern maize breeding (Hallauer and Carena 2009; Hallauer et al. 2010a). The maize ears shows exhibitions or maize ear competitions with scorecards became very popular in USA. The scorecard was an idealized list of what a good maize ear should look like and corresponded to a combination of characters. As an example, Iowa corn growers' association defined a score card in which they punctuated general appearance (25 points divided by ear size and shape, filling of butts and tips, straightness of rows, kernels, uniformity), productiveness (60 points divided by maturity, vitality and shelling percentage) and breed type (15 points divided by size and shape of the ear and dent of kernel, grain and cob color and arrangement of rows). A similar score card was established by the Illinois corn Breeder association in 1890 with the purpose of "developing an interest in better seed corn" (Klesselbach 1922; Winter 1925; Fitzgerald 1993; Hallauer et al. 2010a). This combination of traits allowed to set maize ears ideotypes, which gradually changed the selection procedures used by farmers, producers and breeders (Bowman and Crossley 1908) contributing

for the development of better performing cultivars (*e.g.*, 'Reid yellow dent'). This selection was based on single ideotypes depending on a personal concept and success relied on the patience and perseverance of the person performing the selection (Hallauer and Carena 2009; Hallauer et al. 2010a).

Specifically, different selection paths could lead to the same results on yield comparative tests; *e.g.*, the 'Krug' maize population, that was not selected to meet score card standards, yielded similarly to the 'Reid yellow dent' maize population that was selected according with score cards (Hallauer et al. 2010a).

In Portugal, a maize show was initiated in 1992 at Paredes city. The regional "Sousa Valley Best Ear" competition started as a local and amateur initiative with the purpose of electing the best maize ears within the Sousa Valley Region. The "Sousa Valley Best Ear" competition, is still active nowadays due to its recognition by the community. It tracks interesting germplasm and proactive farmers, promotes rural human development on both anthropological and sociological aspects and its ear value formula is a pedagogic tool for farmers by providing information on relevant traits to be considered for ear evaluation and, indirectly, for breeding selection.

This region is one of the most important Portuguese maize production areas, where traditional maize varieties with technological ability for bread production are still currently produced and improved by farmers, representing a rich source of interesting traits and germplasm for modern maize breeding.

VI.2.2 Questions, motivations and applications

According to the best of our knowledge, there are no reports on studies to select the best formula that relates maize ears traits, the most popular farmer evaluation approach, with the measured yield with such specificities as to be used with the extended objectives of the Sousa Valley best ear' competition. However selection indexes since its introduction (Smith 1936) and development (Williams 1962; Lin 1978; Baker 1986) have been routinely used by breeders where selection is influenced by the relative weight they give to each trait. Visual acuity and experience fine-tune their final decisions. In this sense plant breeding has been considered an art rather than a scientific method (Hallauer et al. 2010a).

Initially, the evaluation of the "Sousa Valley Best Ear" was based on the total number of kernels per ear. However, the maximum number of kernels per ear could be found in a popcorn ear (*e.g.* 164 g for popcorn *versus* flint 'Pigarro' with 345 g for thousand kernel weight), presenting smaller kernels, and meaning that the competition could be won by small ears against larger ears with larger kernel sizes, but smaller number of kernels (Moreira 2006). To solve this ear value problem, an empirical formula to be used on the following editions was developed by Silas Pêgo, a Portuguese maize breeder, specialist in participatory breeding approaches (Moreira 2006). With this formula Silas Pêgo saw an opportunity not only to fulfill the initial function of the competition (*i.e.* to select the best maize ear based on

the ear grain yield prediction on the kernel weight at 15% moisture), but also to advice farmers about selection or traits that could be used to improve yield.

Silas Pêgo's ear value formula (EV formula) was defined as:

$$\mathrm{E}V = (0.6 \times KW + 0.2 \times \mathrm{L} + 0.15 \times \mathrm{R} + 0.05 \times \mathrm{KN})/4$$

Equation VI.1

where **KW** stands for kernel weight (grams) at 15% moisture, **L** for ear length (centimeters), **R** for kernel row number and **KN** for total number of kernels.

The traits included in the formula, and their respective coefficients, were selected based on published correlations with yield (Hallauer et al. 2010a). Exception was the number of kernels that was kept for historical reasons, since it was the first trait to be evaluated on the 1st year competition. In particular, the kernel weight at 15% moisture was chosen because it expresses directly the ear grain yield (the most important yield trait) and has a genetic correlation of 0.25 with yield (Hallauer et al. 2010a). The ear length and kernel row number were also chosen due to their established positive genetic correlations with yield (0.38 and 0.25 respectively) (Hallauer et al. 2010a). However, despite its superiority among the genetic correlations, it is known that the ear length is not successfully used in indirect selection to increase grain yield (Hallauer et al. 2010b). This can be explained by the lack of proper allele combinations, so as by the low heritability and epistatic genetic correlations with other traits (Hallauer et al. 2010a). In this way, its attributed coefficient was only

of 0.2 and subsequently a smaller 0.15 was attributed to kernel row number taking into consideration the respective correlations with yield. However, ear length and kernel row number are negatively correlated (−0.16). In this way, maximization of both traits, by selecting longer ears and higher kernel row numbers, would emphasize the ear fasciation trait expression. Fasciation describes the enlargement of the plant apex by unregulated proliferative growth (Jones 1935; Taguchi-Shiobara et al. 2001; Busch, Benfey 2010) and is normally characterized by abnormal flatten ear types with higher kernel row number (Pêgo, Hallauer 1984). These traits are still highly important to Portuguese farmers in traditional agricultural systems. Indeed during a collecting mission that took place in 2005 (Vaz Patto et al. 2007), 56% of the collected traditional maize landraces had some degree of fasciation *versus* the 10% observed during the 1980's collecting missions. This trait is quite important on agricultural systems requiring a certain level of plant plasticity, such as traditional systems. In fact, we observed that the expression of ear fasciation varies with the environment (Mendes-Moreira et al. 2008) suggesting that plants could regulate their fasciation expression, according to plant density, production system, availability of nutrients or other external factors.

The main aims of this study were to test if the developed EV formula is the best ear ranking option, and in case there is space for improvement, implement an upgraded new formula by regression analysis. For that analysis we used not only data measured on the

ears (the traits most commonly used by farmers) but also data collected from the corresponding field trials (such as yield), from different maize populations usual winners of the "Sousa Valley Best Ear" competition.

Due to the nature and objectives of the ear competition, the method used to select the variables explaining yield must be fully interpretable. The ear rankings based on the yield estimations/ predictions obtained by the ear value formulas must be understood by the farmers and they should know to which extent each variable affects the final ranking. Multivariate regression and multivariate adaptive regression splines (MARS) are two known fully interpretable regression methods. Multivariate regression is a well-known statistical method for regression (Kleinbaum et al. 2008.). It is parametric in the sense that requires certain assumptions in order to best fit the model to the given data. By contrast, MARS (Friedman 1991) is non-parametric, and consequently there are no assumptions for its application.

After the development of the new potential ear ranking models (new EV formulas), the goodness of such ranks should be evaluated according to its adequacy to the true yield.

Research in information retrieval and data mining can contribute to solve this ranking question. Instance ranking is a sub-area of preference learning (Fürnkranz, Hüllermeier 2011). However, instance ranking measures are also used in the evaluation of document relevance, a relevant topic in the area of information

retrieval (Manning et al. 2008). Generalizing, the aim is to rank a given set of instances according to their expected values.

The existing instance ranking evaluation measures aim to evaluate the quality of a given rank by comparison against a given ground true rank. One of the most used evaluation measures for instance ranking is the normalized discount cumulative gain (NDCG) (Järvelin and Kekäläinen 2000). The prefix R. is used to emphasize the recall-based nature of this measure. Recall is an evaluation measure used mainly in classification problems. It measures the fitness of a test in detecting the positives. Recall is also known as sensitivity. In our case study, the ideal rank is the yield rank ordered by decreasing order of the yields values.

The R.NDCG measure has several merits. It deals with any number of different values in the variables used for ranking. In our case, these variables (the yield and the predicted yield) are both continuous and consequently both of them potentially have a large number of different values.

It gives higher weights to higher positions in the rank. This is important because in the maize ear competition, the relevance of the right position is more critical in the top of the rank and also because, the farmers that select for next sowing season will select the maize ears on the top of the rank.

It is possible to limit the R.NDCG measure to the k top values due to the cumulative nature of the measure. Once again, this fits well with the purpose of the ear competition.

Although R.NDCG penalizes top ranked yields that have bottom ranked predictions, it does not penalize top ranked predictions that have bottom ranked yields. An example of this situation is an ear with yield and predicted yield of 10Mgha−1 and7Mgha−1, respectively *versus* another ear with respective yield and predicted yield of 7Mgha−1 and 10Mgha−1. In this example, the R.NDCG measure will penalize only the first case. For maize ear competitions, both situations should be penalized.

While R.NDCG can be seen as a recall measure, we need a measure that evaluates both precision and recall. Due to the exposed, we conclude that none of the existing ranking measures of regression models is applicable to the particular context of the ear ranking at the "Sousa Valley Best Ear" competition, where both precision and recall are needed.

In this way our particular goals were then: (1) to test alternative interpretable regression methods (namely from multiple linear regression and multiple adaptive regression splines) to provide new ear value formulas that better estimates the yield potential using ear traits; (2) to develop a new instance ranking method, allowing to select the best new ear value formula to be used on the ear competition; (3) to identify a set of traits that will help farmers on selection toward better yield; and (4) to compare the ranking results obtained by the original EV formula and the newly one developed, using data from the "Sousa Valley Best Ear" competition.

VI.3. Results

VI.3.1 Alternative interpretable regression methods to provide new ear value formulas that better estimates yield potential

The method used for testing alternative interpretable regression methods randomly splitted the given instances into 10 subsets of equal size (35 instances in each). The model was subsequently trained using the instances from 9 subsets, and tested in the remainder subset. This process was repeated 10 times always leaving a different subset for testing. Ten different formulas were generated from each of the 10 methods that learn the models/formulas from data (Table VI.1), *i.e.*, all except EV, with the indication of what were the major traits related with yield potential in each of them. The frequency of the presence of each trait in the ten formulas (Table VI.2) indicated plant stand as an important trait for models including field variables. Ear weight was another example of a trait constantly represented is the several models.

Table VI.1 Ranking of the 11 regression methods using different measures for ranking evaluation.

Name	varIndex			R.NDCG			P.NDCG			PR.NDCG			*Stat.V*[a]
	Eq. (4)			Eq. (1)			Eq. (2)			Eq. (3)			
EV				0.8174873	(8.00)	*11*	0.8990219	(4.07)	*2*	0.8582546	(5.93)	*7*	abc
mlr.varsEV	0.1450455	(8.57)	*10*	0.8273811	(5.14)	*4*	0.8974835	(4.86)	*4*	0.8624323	(4.86)	*3*	abc
mlr.varsEVeKD	0.1403184	(7.93)	*9*	0.8286758	(4.36)	*2*	0.8764753	(9.36)	*11*	0.8525755	(8.43)	*10*	bc
mlr.ear	0.1232281	(3.57)	*3*	0.8289659	(5.29)	*5*	0.8911580	(6.00)	*7*	0.8600620	(5.71)	*6*	abc
mlr.ear.best4	0.1274515	(4.50)	*4*	0.8297761	(4.86)	*3*	0.8921803	(5.57)	*5*	0.8609782	(4.64)	*2*	ab
mlr.all	0.1140223	(1.71)	*1*	0.8197224	(6.57)	*7*	0.8926395	(5.57)	*6*	0.856181	(6.57)	*8*	bc
mlr.all.best4	0.1215690	(3.00)	*2*	0.8302934	(4.29)	*1*	0.8886621	(7.36)	*8*	0.8594778	(5.50)	*5*	abc
mars.ear	0.1336945	(6.07)	*6*	0.8208717	(7.43)	*10*	0.9233143	(1.29)	*1*	0.8720930	(2.29)	*1*	a
mars.ear.best4	0.1394982	(7.71)	*8*	0.8246521	(6.71)	*8*	0.8807814	(8.14)	*9*	0.8527168	(8.00)	*9*	bc
mars.all	0.1264123	(4.64)	*5*	0.8228027	(6.29)	*6*	0.8751026	(9.29)	*10*	0.8489527	(8.71)	*11*	c
mars.all.best4	0.1387509	(7.29)	*7*	0.8212064	(7.07)	*9*	0.8971842	(4.50)	*3*	0.8591953	(5.36)	*4*	abc

In brackets, the average rank for the 14 groups obtained for each of the eleven methods. In italic the relative ranking based on the previous average rank's

a - Stat.V. - statistical validation of PR.NDCG results for the post-hoc Nemenyi test were obtained after the Friedman rank test. The Friedman rank test rejects the null hypothesis with a p-value < 5%. The post-hoc Nemenyi test was used to validate whether the difference of the averaged ranks is larger enough to be statistically valid for the desired level of significance. Using this test, two methods are statistically different when the difference of the averaged ranks is larger than 4.12. Significant differences exist if no common letters exist among groups.

Table VI.2 Frequency of the presence of each trait in the ten models resulted from the respective iterations. Frequency varies from zero to one.

Measurements	mlr.ear	mlr.ear.best4	mlr.all	mlr.all.best4	mars.ear	mars.ear.best4	mars.all	mars.all.best4
Moisture %	0	0	0.5	0	0	0	0.2	0
Stand x 1000	0	0	0.6	0.6	0	0	0.7	0.7
Cob/Ear weight, Harvest	0	0	0.4	0	0	0	0.2	0.1
Root lodging %	0	0	0.6	0.1	0	0	0.1	0
Stalk lodging %	0	0	0.3	0.2	0	0	0	0
Plant height, cm	0	0	0.8	0	0	0	0	0
Ear height, cm	0	0	0.3	0	0	0	0.2	0.1
Ear Length, cm	0.6	0.2	0.6	0.2	0	0.1	0.1	0
Ear Diameter 1, cm	0.2	0	0	0	0	0	0	0
Ear Diameter 2, cm	0.4	0.3	0.4	0.2	0.1	0	0.4	0
Ear Diameter 3, cm	0.1	0	0.3	0	0	0	0.1	0
Ear Diameter 4, cm	0.1	0	0.3	0	0	0	0.2	0
Kernel-row number 1, nº	0.4	0.1	0.7	0.2	0.2	0	0.4	0
Kernel-row number 2, nº	0.4	0.1	0.7	0.2	0.2	0	0.4	0
Fasciation	0.5	0	0.7	0	0.2	0.1	0	0
Ear weight, g 15% moisture	1	1	1	1	0.9	0.9	1	1
Kernel weight, g 15% moisture	0.2	0.1	0.2	0.1	0.2	0.3	0.2	0
Cob/Ear weight	0.4	0	0.3	0	0	0.1	0.1	0
Ear% Moisture	0.1	0	0.2	0	0	0.1	0.3	0.1
Kernel depth, cm	0.5	0.4	0.5	0.4	0.5	0.7	0.4	0.3
Kernel number, nº	0.5	0.2	0.5	0.1	0.1	0	0.2	0
Thousand kernel weight, g	0.6	0.4	0.6	0.4	0	0.1	0.2	0
Kernel per row, nº	0.3	0.1	0.5	0.1	0.1	0	0	0.1
Cob diameter 1, cm	0.1	0	0.1	0	0	0	0.1	0
Cob diameter 2, cm	0.3	0	0	0	0.1	0	0.1	0
Cob diameter 3, cm	0.3	0	0.4	0	0.1	0.1	0.2	0.2
Cob diameter 4, cm	0.7	0.4	0.5	0.1	0	0	0	0
Medulla 1, cm	0.3	0.1	0.3	0.1	0	0	0	0.1
Medulla 2, cm	0.4	0.1	0.2	0	0	0	0.3	0
Rachis 1 , cm	0.2	0	0.3	0	0.3	0.3	0.5	0.2
Rachis 2, cm	0.3	0.2	0.4	0.2	0.3	0.2	0.5	0.4

VI.3.2 A new instance ranking method, to select the best new ear value formula for ear competition

The proposed precision recall normalized discount cumulative gain (PR.NDCG) is a measure that evaluates both precision and recall, whether the top ranked yields are in the top of the predictions rank, and whether the top ranked predictions are in the top of the yields rank. Let P.NDCG be defined as follows:

Let P.NDCG be defined as follows:

P.NDCG[i]=DCG.ideal[i]/DCG.eval[i]

Equation VI.2

Where *i*, as in Equation VI.15, gives the position until where the ranking is evaluated.

Then, PR.NDCG is defined as follows:

PR.NDCG[i]:=α x R.NDCG[i] + (1-α) x P.NDCG[i], where $\alpha \in [0,1]$

Equation VI.3

The PR.NDCG can be seen as a generic function having as particular case both the recall measure R.NDCG, when ¸= 1, and the precision measure P.NDCG, when α=0.

For the problem of maize ear competitions we propose the use of α= 0.5 because both recall and precision are equally important in this context.

As far as we know, this is a novel measure. Yet, other measures exist combining both recall and precision measures. The usefulness of such

combined measures is discussed by Järvelin and Kekäläinen (2002). However, these authors do not present specific approaches to deal with it. This is done by (Kazai, Lalmas 2006). They present a measure named Effort Precision–Generalized Recall – EP/GR that, applied to the NDCG measure would be :

$$EP/GR = P.NDCG/R.NDCG$$

Equation VI.4

This measure evaluates the gain of the precision measure over the recall measure. None the less this is not useful for the ear ranking problem because the desired evaluation measure should equally weigh both measures instead of evaluating how much the precision surpasses in percentage the recall, as is done by the EP/GR measure. We observed that the variance of P.NDCG is larger in comparison with R.NDCG. For this reason there is a higher correlation between the rankings obtained with P.NDCG and PR.NDCG than the ones between R.NDCG and PR.NDCG (Table IV.1).

Another method that combines both precision and recall measures is the expected precision-recall with user modeling – EPRUM (Piwowarski, Dupret 2006). However, this measure was designed for evaluation of ranks of XML (Extensible Markup Language) objects. This specific problem must deal with the aggregated nature of XML objects. For instance, the relevance of a subsection object can depend on the relevance of its parent section. This characteristic of

XML objects is not meaningful for ear ranks. For this reason, the EPRUM measure is also not applicable to our problem.

The PR.NDCG measure we propose is a linear combination of a precision and a recall measures that, for our problem, is certainly more adequate.

A statistically validated comparison between the eleven methods using the PR.NDCG measure was developed for ranking evaluation (Table 4, the average ranks obtained are in brackets and, in italics, the ranks based on the average ranks). To proceed with the statistical validation we have obtained a p-value for the null hypothesis of equivalence among the eleven ranking methods as 0.00001369 < 0.01, the assumed level of significance. For the post hoc Nemenyi test, the difference of the averaged ranks was larger enough to be statistically valid when the difference of the averaged ranks was larger than 4.12.

The statistical validation indicated that mars.ear was significantly better than mlr.all, mars.ear.best4, mlr.varsEVeKD and mars.all. In addition, mars.all was significantly worse than mars.ear and mlr.ear.best4. In was also registered that on an overall comparison no significant differences among the seven high ranked methods existed.

Nevertheless, based on a complexity reduction priority, a new proposed formula to use in the competition was selected, mlr.varsEV, and named ear value adjusted formula (EVA formula). The EVA formula (Equation VI.6) uses the same variables as the EV formula.

The only differences are the βi coefficients (i=1, …, 4) of the linear formulation.

mlr.varsEV= -7.030877 +0.031605 x KW+0.387825 x L+ 0.337015 x R12 -0.008875 x KN

Equation VI.5

EVA =mlr.varsEV

Equation VI.6

VI.3.3 A set of traits to help farmers on selection toward better yield

The first six ranked methods, excluding mlr.varsEV with fixed traits, in all the ten iterations of each method, consisted of 11 (in mars.all.best4) to 23 traits (in mlr.ear). On these six ranked methods, excluding fixed traits methods, ear weight, kernel depth and rachis 2 were always present (Table VI.2). The cob and ear diameters and kernel per row were also present, but with a more residual contribution. For methods that included field traits (mars.all.best4, mlr.all.best4), plant stand was also a very important trait. However, field traits were not considered in the first three methods and they did not contribute significantly to improve the model when considered.

To select a new ear value formula that better estimated the yield potential of each ear and that could be used for competition, we have used all the 350 instances. Theoretically, this selected model should be more stable than any of the 10 models obtained through the 10-fold cross-validation process (Table VI.2).When comparing the best three ranked methods (mars.ear, mlr.ear.best4 and mlr.varsEV,

i.e., Equation of Table VI.3, Equation VI.7 and Equation VI.5, respectively):

mlr.ear.best4= 4.010951 -26.369230 x CW_EW2 +0.003765 x SW -0.191631 x KW +0.188808 x EW

Equation VI.7

mlr.varsEV= -7.030877 +0.031605 x KW+0.387825 x L+ 0.337015 x R12 -0.008875 x KN

Equation VI.8

it was observable that kernel weight was common to all of them, while thousand kernel weight and ear weight were common only to mars.ear and mlr.ear.best4. Moreover, the mars.ear model (composed by 12 of the possible 24 traits used in the 14 terms of the formula) (Table VI.3) had three of the four traits of mlr.ear.best4 (the exception was the ratio between cob and ear trait). In addition, when comparing mars.ear *versus* mlr.varsEV model we could observe that two traits out of the four were common to both models (length and kernel number where not common). Furthermore, different non common traits can be highly correlated. As an example, the ear length with both ear weight (0.859) and thousand kernel weight (0.770), and also kernel number with ear weight (0.902), indicate that mars.ear and mlr.ear.best4 (with three common traits) *versus* mlr.varsEV methods can be more similar than expected. Therefore, different combinations of formula variables (traits) can have the same effect when yield is addressed, due to the existence of highly correlated variables, *i.e.*, indicating that different models can perform more similarly than expected.

Table VI.3 The mars.ear model using 350 instances, where 14 Terms plus A0 constant were selected from 116 Terms. 12 traits from 24 were selected.

Model	Basis functions
A0 -0.0498379*A1 +14.4230551*A2 -1.3567641*A3 +0.0104145*A4 -0.3503811*A5 +0.0396746*A6 -0.0003114*A7 + 0.0135842*A8 -0.0049833*A9 -0.0142137*A10 -0.0444110*A11 +0.0020012*A12 +0.0395830*A13 +0.0015190*A14	
	A0=Constant= 8.0186081
	A1=max{0,222.336-EW15}
	A2=max{0,KD-1.0475}
	A3=max{0,1.7035-M1}
	A4=max{0,(17.6-R12)}* max{0, (KN-490.551)}
	A5=max{0,(329.723-EW)}* max{0, (KD-1.0475)}
	A6=max{0,(EW-222.336)}* max{0, (Rq1-2.8085)}
	A7=max{0,(KN-490.551)}* max{0, (SW-377.424)}
	A8=max{0,(490.551-KN)}* max{0, (3.425-CD3)}
	A9=max{0,(5.67-ED1)}* max{0, (19.5-R12)}* max{0, (KN-490.551)}
	A10=max{0,(2-Fa)}* max{0, (EW-222.336)}* max{0, (15.53-E.Moisture)}
	A11=max{0,(329.723-EW)}* max{0, (KW-262.833)}* max{0, (KD-1.0475)}
	A12=max{0,(329.723-EW)}* max{0, (262.833-KW)}* max{0, (KD-1.0475)}
	A13=max{0,(329.723-EW)}* max{0, (17.09-E.Moisture)}* max{0, (KD-1.0475)}
	A14=max{0,(EW15-222.336)}* max{0, (SW15-393.221)}* max{0, (2.8085-Rq1)}

VI.3.4 Ranking comparisons between the original EV formula and the newly developed one

Yellow dent and white flint ear data from previous editions of the "Sousa Valley Best Ear" competition were used for comparing the ranks obtained using the new EVA formula and the original EV formula. From this comparison, we observed that 4 ranks (1, 2, 3 and

4) using the EV formula were maintained on the top 5 positions on EVA formula for white flint maize. In the case of yellow dent maize, the ranks 1, 2, 3 and 5 using the EV formula were maintained in the top 10 of the EVA formula rank positions (data not shown). These data indicated that EVA is more independent from kernel weight than EV. Indeed, the EV data rank is almost coincident with the kernel weight rank. In addition when the 350 instances are used we obtain only two top 10 ranks for EV *versus* four in the case of EVA when we consider yield. All these facts indicate EVA as amore refined formula to be used for ear competition and for farmers' selection and evaluation.

VI.4. Discussion

The present work allowed the development of a more effective indicator of maize yield based on ear traits (new ear value formula) that can be used for maize ear competition, but it also allowed to the identification of the most informative traits that can be used by farmers in maize selection.

To start with, the choice of the most appropriate evaluation measures to rank the alternative regression methods was required. The newly developed PR.NDCG ranking method has proven to be the most appropriate method because it combines both recall and precision measures. We believe that PR.NDCG can be also valuable for other problems besides ear ranking. Indeed, in problems where the goal is to know the ranking position according to a certain

forecasted numerical measure, this approach can also be useful. Especially when both a bad instance being well ranked and a good instance being badly ranked are undesirable. An example of such problems could be ranking people according to their potential interest for a company in a human resources selection procedure.

Field traits were not present in the first three ranked interpretable regression methods for yield potential estimation, not having a key role to improve the models. This means that field traits increased measurements costs, but do not improve the yield potential estimation. For this reason we decided to focus specifically on the three first methods. First the ear weight, kernel depth and rachis 2 and second the ear diameters and kernel per row represented not only some of the traits that will help farmers on selection toward better yield, but also some of the eligible variables that can be selected for the ear value formula.

The correlations between ear traits and yield obtained by us indicated a similar tendency to those previously reported (Hallauer et al. 2010a), namely, kernel weight (0.78; 0.25), ear length (0.69; 0.38), kernel-row number (0.12 with R1 = 0.09 and R2 = 0.14; 0.24), kernel number (0.69; –) and kernel depth (0.53; 0.51).

The final selection of a reduced number of variables from the initial set can in part be explained by the high correlation observed among them (*e.g.* correlation between ear length and ear and kernel weight, 0.859 and 0.858 respectively).

Similarities among methods, without significant differences in their ability to rank, led to selecting the mlr.varsEV model (Equation VI.5) as the best method for ear competition and farmers use, because it is much less complex than mars.ear and because it does not include field variables reducing the associated measurement costs. In addition, mlr.varsEV has the same variables than EV (Equation VI.1), which are already familiar to the farmers. Indeed, based on our present results, the original set of variables proposed by Silas Pêgo seems to be the most promising, after a required adjustment of their coefficients.

In comparison with scorecards with a range from 0 to 100%, the values obtained with EV and EVA have an open scale, which allows the comparison along time. In the white flint category of Sousa Valley competition (associated to maize with bread technological ability), 185 ears competed during 13 years. From this sample, a minimum was registered in 1999 with an EVA of 4.28, corresponding to an ear with 153.91 g/ear at 15% moisture, 15.7 cm length, 22 row number and 765 kernels. The maximum EVA for white maize flint was 16.41 (obtained in a ear competing in the year 2000) with 396.21 g/ear at 15% moisture, 27.30 cm for length, 24 row number and 874 kernels per ear. This means that for this maximum EVA value and for a plant stand of 70,000, the potential yield would be 27.73 Mg (if all the plants with one ear with 396.21 g). In this same germplasm type the average value of EVA was 10.25, which represented an average value of 256.09 g/ear at 15% moisture 22.16 cm for length,

20rownumberand 696 kernels per ear. During the white maize flint competition the maximum values observed for an ear were 1136 kernels per ear, a maximum of 29 cm length and 40 rows (indicating the presence of fasciation).

For the yellow dent category of Sousa Valley competition (majority of hybrids) the minimum EVA (4.61) was obtained in 1998. This minimum represented an ear of 131.11 g of grain 15% moisture, with 19.5 cm length, 14 rows and 539 kernels per ear, with a thousand kernel weight of 243 g. The maximum EVA for this germplasm type, obtained in 1997, was 19.33, and represented an ear of 553.94 g of kernel weight 15% moisture, with a length of 29.9 cm, 18 rows and 992 kernels with a thousand kernel weight of 558 g. This maize winner, as the majority in all the editions in this germplasm type was from 'Fandango', a synthetic population developed in Portugal. The average EVA for yellow dentwas11.79.A simulation for a plant stand of 70 000 plants per hectare (considering one ear per plant), indicated an average yield potential of 22.06 Mg/ha, with a range of 9.18–38.88 Mg/ha maize potential.

To establish an international standard for EVA value comparison, data from an average of 180 ears each for 'BS21', 'BS22', 'TEPR-EC6', *i.e.*, North American standard populations were analyzed. Their EVA values were 4.51, 5.10 and 3.84 with a kernel weight of 135.56, 148.32 and 126.85; 14.6, 16.8 and 14.8 for ear length and 15.5, 15.0 and 14.5 for row number, respectively. Among the Portuguese developed maize population, the average EVA for 'Fandango' was

8.41 (160 ears), for 'Pigarro' mass selection (216 ears) was 6.83, and 6.34 for 'Pigarro' recurrent selection (90 ears).

The empirically derived models in this study were specific to the range of the populations used in the competition of "Sousa Valley Best Ear". To some extent, such models can be calibrated for use with other maize populations, as can be seen for original EV (formula (1)) and EVA (formula (14)).

$$EV = (0.6 \times KW + 0.2 \times \mathrm{L} + 0.15 \times \mathrm{R} + 0.05 \times \mathrm{KN})/4$$

Equation VI.9

EVA =mlr.varsEV= -7.030877 +0.031605 x KW+0.387825 x L+ 0.337015 x R12 -0.008875 x KN

(Equation VI.5 and Equation VI.6)

In this way, this type of formula development could play an important role not only for maize yield prediction, but in any other plant species under participatory plant breeding research, where farmers need the right tools to tackle selection for different species and agroecosystems (Soleri et al. 2000; Soleri, Cleveland 2009; Machado et al. 2011; Dawson et al. 2013).

This is, from the best of our knowledge, the first study where this particular mathematical regression approaches were used for fine tune maize yield potential estimators based on ear traits. Nevertheless these mathematical approaches have recently started to be more often applied to answer biological questions. Also in the case of maize, multivariate adaptive regression splines have been

used to identify relationships between soil and maize production properties, helping to decipher potential cause and effect processes, concluding that soil physical parameters were more important than nutrients for maize yield estimation (Turpin et al. 2005). On sweet potato (*Ipomoea batatas*), linear regression and data mining methods were used to improve the prediction of sweet potato harvest on the basis of agrometeorological variables (Villordon et al. 2009). With more broad application objectives, similar approaches using regression tree analyses, bagging trees, random forests and multivariate adaptive regression splines, were evaluated and compared for predictive vegetation mapping under current and future climate scenarios to model the distribution of a large number of tree species under climate changes scenario. Excluding multivariate adaptive regression splines, it was suggested to combine the other methods because they provided a means to accurately map organism distributions and a mechanism that provided a better understanding of the drivers of present and future distributions (Prasad et al. 2006).

VI.5. Conclusions

Yield is an expression of fitness and radical changes in one yield component are accompanied by adjustments in other component(s), implying the existence of correlated changes of gene frequencies (Hallauer, Carena 2009). This fact explains that the same yield increase can be obtained by selecting for different trait combinations

originating different phenotypes (*e.g.* bigger ear *versus* prolificacy, prolificacy *versus* higher densities). These different phenotypes can be better adapted to particular systems, such as higher density and smaller ears for intensive systems or lower densities with bigger ears for traditional systems (*e.g.* intercropping systems).

The data set of the 350 instances, not being a group of the best ears, but a set of populations where the best Sousa Valley Ears come from, represents a broad range of plants and ears. This broad range of data highlighted some of the traits that can be used both for selection *per se* and selection for best maize ear competition goals. As an example, ear weight, kernel depth and rachis 2 were considered of major importance according to our results from 10 models obtained throughout the 10-fold cross-validation process, followed by cob and ear diameters and kernel per row. Stand was the most important field variable (but not used for maize competition). These data were obtained from the representativeness of the traits regarding the six best ranked methods, excluding fixed traits models. With exception of the first method (mars.ears), composed by 12 of the possible 23 traits, the following four ranked methods obtained had only 4 variables or terms.

The selected EVA formula showed to be the best compromise solution due to: less complexity than the first models and no inclusion of field traits which proved to be nonessential (models with field traits were classified in 4th and 5th rank), *i.e.*, field traits did not contribute to improve the rank, and high correlations existed among

the majority of the variables that were not common among the best three models (*e.g.* mars.ear, mlr.ear.best4 and mlr.varsEV). The EVA traits formula indicates that kernel weight, ear length, kernel row number and number of kernels are some of the most important traits both for selection and for the best ear competition. Due to its simplicity, EVA formula can be easily adopted by farmers and by local associations interested in germplasm conservation and development. The use of this EVA formula on the maize ear competition provides the missing link between the farmer and the community to engage both on collaborative research. Additionally, a smaller number of traits are less expensive to measure. Consequently it can be used as a tool in participatory plant breeding (PPB) projects where quantitative information is collected by farmers in order to improve their own selection procedures. Furthermore they can be expanded to pre-breeding, on-farm conservation, adaptation to organic, low input agriculture or climate changes, or to better understand breeding selection procedures evolution along time and from breeder to breeder.

In conclusion, the EVA formula is a starting point for a long term engagement with germplasm development and an open door to a better understanding of quantitative genetics by farmers.

VI.6. Material and Methods

VI.6.1 Plant material and traits information

A. For the development of the new EV formula, we have used simultaneously plant field trial data and ear traits from three different maize populations, 'Pigarro', 'Fandango' and 'Nutica'. 'Pigarro' is a white flint improved landrace (Mendes-Moreira et al. 2008) and both 'Fandango' and 'Nutica' are yellow dent synthetics (Mendes- Moreira et al. 2009). The 'Pigarro' and the 'Fandango' populations were chosen because they represent the usual winners of the "Sousa Valley Best Ear" competition during the past 13 years. The 'Nutica' population, is the ancestor of the 'Fandango' population, and was included on this study to provide more resilience to the ear value model due to its higher diversity.

The 'Pigarro' and 'Fandango' maize populations are being selected under a long run participatory plant breeding project since 1984 (VASO Project). Seeds from each of their farmers' mass selection and breeders' recurrent selection cycles are always kept on cold storage.

The 'Pigarro', 'Fandango' and 'Nutica' multi-location field trials were established on the Portuguese locations of Montemoro- Velho, Coimbra and Lousada, during one to four years (Mendes-Moreira et al. 2008, 2009). These trials included seven to eight farmer mass selection cycles of 'Fandango', considering 'Nutica' as the initial selection cycle, and from now on named 'Fandango + Nutica' cycles, and eight farmer mass selection cycles plus three recurrent selection cycles of 'Pigarro' (Table VI.4). Field trials were established on a plot

basis, each plot corresponding to 9.6m2 (2 rows, 6.4m long with 0.75m between rows), with 3 repetitions (Table VI.4). Trait measurements were obtained per plot, as in the case of yield and the root and stalk lodging, or based on 20 individual measurements, as in the case of the ears traits or plant height (Table VI.5). A total of 32 traits (8 field traits and 24 ear traits) were measured and used in the data analysis (Table VI.5).

Data were obtained from 305 instances from 'Pigarro' and 175 instances from 'Fandango + Nutica'. These instances (also named tuples, records or examples) correspond to a set of traits with one measured value (*e.g.* yield, root and stalk lodging) or the average of 20 measurements (*e.g.* ears traits) per plot. In the case of 'Fandango + Nutica', 7, 5 and 8 selection cycles were tested respectively in 2005, 2007 and 2008, in three locations, with three repetitions each, *i.e.*, 180 instances. From these 180 instances, five were incomplete and rejected, for this reason we have considered only 175 instances. In the case of 'Pigarro', 10 cycles of selection were tested in 2005, 2006 and 2007, 11 in 2008, at three locations for the odd years and two in the even years, with three repetitions per location. This resulted in a total of 306 instances. From these a sample of 175 instances was randomly selected to build a balanced sample of 350 instances, together with the 175 'Fandango + Nutica' instances previously obtained.

B. For the comparison of the ranking results of the original and the newly developed ear value formulas, we used ear data obtained from

the records of the “Sousa Valley Best Ear” competitions previous editions.

The “Sousa Valley Best Ear” competition started in 1992, and since then every year, ear data is collected by the evaluating organizations NUMI (maize breeding station, Braga, Portugal) or DRA (Regional Agricultural Services). To enter the competition, farmers should deliver and register their best ears at Cooperativa Agrícola de Paredes (Paredes Farmers Association) and ears are scored using the EV formula (Equation VI.1), previously described. In this way, the evaluated traits are: kernel weight at 15% moisture (KW), ear length (L), kernel row number (R) and number of kernels (KN). Ear types are ranked separately in four categories (by color: yellow and white maize and kernel type: dent and flint), each category with its own winner. During the evaluation process the ears are photographed and four kernel ear rows removed. Two of them are used for moisture content determination and the other two rows are sent to the BPGV (Portuguese Plant Germplasm Bank) in order to keep a sample of this germplasm in cold storage. Since 1992 and during the following 13 years of ear competition, 454 yellow dent and 185 white flint ears were evaluated. These are the data now used for comparing the ranking results of the ear value formulas.

VI.6.2 Statistical methods

VI.6.2.1 Interpretable regression methods

Multivariate regression and multivariate adaptive regression splines (MARS) are two known fully interpretable regression methods. Multivariate regression tries to fit a given formula (linear, polynomial, exponential, or other) to the data while MARS assumes a linear formula, but fits it locally. Multivariate regression is not sensitive to over fitting while MARS is.

In this study, from the existing formulations of multivariate regression, we will use the linear one, called multivariate linear regression (MLR).

Table VI.4 Maize populations' characterization: kernel type, number of instances per population, origin and references.

Population	Kernel type	Data[a]	Origin	Area/plot, ears/plot	References
Nutica	Yellow dent	18	Synthetic Pt (80% USA germplasm)	9,6 m^2; 20 ears /plot	Moreira et al, 2009
Fandango	Yellow dent	157	Synthetic Pt (80% USA germplasm)	9,6 m^2; 20 ears /plot	Mendes-Moreira et al, 2009
Pigarro	White flint	305	Populations, Pt	9,6 m^2; 20 ears /plot	Mendes-Moreira et al, 2008

[a] The number of instances per population corresponds to the product of selection cycles (for 'Pigarro' and 'Fandango'), years, locations with three reps, with exclusion of instances that not have the complete set of traits.

Table VI.5Traits measured per location and per plot, codes, and respective units of measure or scales. Means and Standard Deviation of the populations Fandango, Nutica and Pigarro. Correlation between yield and other traits.

	Data/ Plot	Codes	Scale/ units	Fandango Mean		Std. Dev.	Nutica Mean		Std. Dev.	Pigarro Mean		Std. Dev.	Correlations with Yield
Yield*	1	Yield	Mg ha-1	8.27	±	1.82	8.98	±	1.84	6.90	±	1.69	1.00
Moisture at harvest*	1	%Moist.	%	28.01	±	6.37	30.07	±	3.94	28.11	±	4.00	0.00
Stand*	1	Plants ha^{-1}	ha^{-1}	48254.58	±	4468.68	48621.89	±	2804.27	49399.71	±	3185.05	0.07
Harvest. Cob/Ear weight*	1	CW/EW		0.22	±	0.05	0.20	±	0.03	0.24	±	0.05	-0.04
Root lodging*	1	R	%	0.04	±	0.06	0.03	±	0.04	0.03	±	0.04	0.07
Stalk lodging*	1	S	%	0.06	±	0.10	0.05	±	0.07	0.07	±	0.06	-0.29
Plant height*	20	H	cm	270.73	±	44.89	258.09	±	35.74	232.61	±	41.37	**0.57**
Ear height*	20	H1E	cm	152.15	±	31.75	140.76	±	29.20	138.01	±	25.51	0.48
Ear Length	20	L	cm	20.73	±	2.46	20.81	±	1.17	17.26	±	1.55	**0.69**
Ear Diameter 1	20	ED1	cm	5.41	±	0.54	5.34	±	0.15	5.91	±	0.46	0.23
Ear Diameter 2	20	ED2	cm	5.15	±	0.49	5.12	±	0.16	5.48	±	0.39	0.37
Ear Diameter 3	20	ED3	cm	4.87	±	0.54	4.72	±	0.18	4.92	±	0.55	0.32
Ear Diameter 4	20	ED4	cm	4.69	±	0.48	4.59	±	0.18	4.41	±	0.33	0.48
Kernel-row number 1, nº	20	R1	nº	16.94	±	2.44	16.64	±	0.70	18.74	±	2.40	0.09
Kernel-row number 2, nº	20	R2	nº	16.73	±	2.36	16.32	±	0.57	17.89	±	2.55	0.14
Fasciation	20	Fa	1 to 9	1.56	±	0.79	1.87	±	0.50	1.65	±	0.94	0.03
Ear weight, g	20	EW	g	270.65	±	46.83	268.66	±	35.53	196.24	±	34.88	**0.80**

	Data/ Plot	Codes	Scale/ units	Fandango Mean		Std. Dev.	Nutica Mean		Std. Dev.	Pigarro Mean		Std. Dev.	Correlations with Yield
Kernel weight, g	20	KW	g	229.50	±	39.50	229.14	±	29.02	163.93	±	28.59	**0.78**
Ear, Cob/Ear weight	20	E CW/EW		0.15	±	0.02	0.15	±	0.01	0.16	±	0.02	0.00
Ear% Moisture	20	E% Moist	%	17,48	±	3.39	17.33	±	2.22	15.95	±	1.38	0.12
Kernel dept	20	KD	cm	1.22	±	0.11	1.24	±	0.05	1.00	±	0.04	**0.53**
Kernel number	20	KN°	n°	618.26	±	93.10	632.86	±	42.71	487.97	±	75.28	**0.69**
Thousand kernel weight	20	SW	g	373.27	±	46.12	365.40	±	25.99	340.00	±	29.53	**0.59**
Kernel per row	20	NR	n°	39.14	±	4.52	40.77	±	2.01	29.03	±	2.34	**0.62**
Cob diameter 1	20	CD1	cm	3.56	±	0.40	3.43	±	0.16	4.41	±	0.41	-0.03
Cob diameter 2	20	CD2	cm	3.21	±	0.32	3.14	±	0.15	3.84	±	0.32	0.04
Cob diameter 3	20	CD3	cm	2.95	±	0.37	2.78	±	0.14	3.49	±	0.53	0.05
Cob diameter 4	20	CD4	cm	2.74	±	0.28	2.64	±	0.11	2.89	±	0.25	0.13
Medulla 1	20	M1	cm	1.45	±	0.29	1.40	±	0.13	2.29	±	0.31	-0.12
Medulla 2	20	M2	cm	1.12	±	0.19	1.13	±	0.09	1.74	±	0.22	-0.11
Rachis 1	20	Rq1	cm	2.61	±	0.33	2.46	±	0.15	3.49	±	0.36	-0.06
Rachis 2	20	Rq2	cm	2.20	±	0.25	2.15	±	0.09	2.90	±	0.29	0.00

***** - Field data, i.e. measured in the plot or at plot level, generally it corresponds to one observation per plot with exception of plant and ear height. The other data are measured at ear level and correspond to twenty observations by plot.

VI.6.2.2 Multivariate linear regression (MLR).

Given a vector $\boldsymbol{x}[1..n]$ of n explanatory variables, and a response variable y, the goal of MLR is to define the linear relation between $\boldsymbol{x}[1..n]$ and y that minimizes a given metric error, typically, the squared error. The relation between $\boldsymbol{x}[1..n]$ and y is given by:

$$y_j = \beta_0 + \sum_{i=1}^{n} \left([\beta]_i \times \boldsymbol{x}_j[i]\right) + E_j$$

Equation VI.10

The $\beta_i (i = 0, \dots, n)$ coefficients are determined in order that the linear model best fits the given data, according to the metric error. The E_j values are the random errors and cannot be determined. Consequently, since the β_i $(i = 0, \dots, n)$ coefficients are already defined, the prediction $\hat{y}_t$ of an unknown y_t for a given $\boldsymbol{x}_t[1..n]$, is obtained as follows:

$$\hat{y}_t = \beta_0 + \sum_{i=1}^{n} \left([\beta]_i \times \boldsymbol{x}_t[i]\right)$$

Equation VI.11

The most important assumptions of MLR in terms of prediction accuracy are:

Linearity: as expected a linear model only gives acceptable results when there is a linear relation between the explanatory variables and the response variable.

Normality: the calculus of the $\beta_i \ (i = 0, \ldots, n)$ coefficients assumes that the errors (predicted minus observed values) are normally distributed. If not, the linear model is not fitted to the given data as well as it could.

Right choice of the explanatory variables: for problems with many explanatory variables, in particular, the choice of the right subset to use must be properly addressed. This is true for any predictive model, although some algorithms embed a method for variable selection. It should be emphasized that there is no good variable selection method that can compensate for a bad definition of the initial set of variables. In our particular case, the initial set of used variables was based on the bibliographic data available and our field experience. However, there is no guarantee that using all these variables is the most advised approach for the prediction task as compared to using only a subset of them. For this reason, a method for variables selection should also be used. MLR does not embed any variable selection method. However, the majority of the statistical software packages include methods for variable selection that can be used together with MLR. Two of such methods are forward and backward selection. Forward selection begins by including the variable that best explains the response variable. This is repeated until the addition of a new attribute does not add value to the fitness of the model to the data. Backward selection begins with all variables and removes the one that least contributes to fitness of the model to

the data. It stops when removing any of the remaining variables does not add value to the goodness of fit.

VI.6.2.3 Multivariate adaptive regression splines (MARS).

MARS is included in the data mining methods and is a highly automated, data-directed, regression analysis tool designed to detect variable interactions and to build functions able to estimate the response of a target variable to multiple predictor variables using the available data.

Indeed, comparing against MLR, MARS is a set of connected linear pieces. The connection knots are obtained using the recursive partitioning algorithm (Breiman et al. 1984). These knots are parameter values representing a functional change in the behavior of the curve. The more knots the curve has, the higher is its adjustment to the training data. Hence, too many knots may result in over fitting data, reducing the ability of this curve to predict the yield of a new ear.

MARS model is expressed as follows:

$$\hat{y} = \beta_0 + \sum_{i=1}^{I} \beta_i \times \prod_{j=1}^{J_i} \begin{cases} \max(o, x[k_{ij}] - c_{ij}) \\ \max(o, c_{ij} - x[k_{ij}]) \end{cases}$$

Equation VI.12

where $\beta_i (i = 0, \cdots, I)$ and $c_{ij} (i = 0, \cdots, I; j = 1, \cdots, J_i)$ are constants determined by MARS.

The number of MARS terms ($I+1$, where the plus 1 refers the β_0 coefficient) is determined empirically. However, a maximum number can be specified. When a term has $J_i > 1$ it means that it models interactions between two or more variables. The maximum degree of interaction can be also specified.

The MARS only assumption, common to all data mining algorithms, is the relevance of the explanatory variables to explain the response variable. MARS uses also the forward and the backward methods for variable selection, described previously for MLR in the end of Section VI.6.2.3

VI.6.2.4 Precision recall normalized discount comulative gain (PR.NDCG): a new instance ranking measure

To learn how to rank, existing instances with the information about the best way to rank them (usually known as ground true value) are used to test the model.

As previously referred, a new ranking measurement that evaluates both precision and recall was needed to rank our regression models.

The normalized discount cumulative gain (NDCG) (Järvelin and Kekäläinen 2000) is one of the most used recall based ranking evaluation measure. This measure compares the discount cumulative gain (DCG) of a given rank (DCG.eval) against the DCG of an ideal rank (DCG.ideal), *i.e.*, the ground true value. In our case study, the instance represents the basic unit, where all the traits we measured and used (Table VI.5) are present. The DCG is a rank that decreases

the weights of the Gain as long as the position in the rank decreases. The level of this decrease depends on the choice of a b value. We have used b = 2 in all experiments because this value decreases more the gain value as long the position goes down in the rank (Järvelin, Kekäläinen 2002). Example: for i=7 and G[i]=5 the results for $\frac{G[i]}{\log_b i}$ would be 1.78 and 2.82 for b equalized to 2 and 3, respectively. Indeed, larger values of b result in smoother discounts in lower rank positions. NDCG is described in Equation VI.15 following closely the original formulation given by Järvelin and Kekäläinen (2000, 2002).

$$CG[i] = \begin{cases} G[1], \text{if } i = 1 \\ \text{CG[i-1]} + \text{G[i]}, \text{otherwise} \end{cases}$$

Equation VI.13

$$DCG[i] = \begin{cases} CG[1], \text{if } i < b \\ \text{DCG[i-1]} + \text{G[i]}/\log_b i, \text{ if } i \geq b \end{cases}$$

Equation VI.14

$$R.NDCG[i] = DCG.eval[i] / DCG.ideal[i]$$

Equation VI.15

i gives the position until where the ranking is evaluated. The prefix R. is used to emphasize the recall-based nature of this measure. In our case study, the ideal rank is the yield rank ordered by decreasing order of the yields values. Similarly, we evaluate the rank yield prediction using the decreasing order of the yield prediction.

It is possible to limit the R.NDCG measure to the k top values due to the cumulative nature of the measure, as already mention in the introduction. Once again, this fits well with the objectives of the ear competition. The k value depends necessarily on the problem. We have used k = 10 because only the top rank ears are meaningful in ears contests.

Nevertheless, while R.NDCG can be seen as a recall measure, we indeed needed a measure that evaluated both precision and recall. Consequently, we proposed a new approach – the precision recall normalized discount cumulative gain (PR.NDCG) for which we present the measure calculation on the results Section VI.3.2.

VI.6.3 Data analysis

Data analysis from field trials was performed in four steps (Figure VI.1): (1) In a first step, yield was predicted using interpretable regression methods, as discussed in Section VI.6.5; (2) in a second step, using the results obtained with the different interpretable regression methods, ears were ranked using different ranking evaluation measures; (3) in a third step, the different interpretable regression methods were statistically compared according to the newly developed PR.NDCG measure; and (4) finally, in order to obtain a more stable model in opposition to the 10 different models obtained by the cross validation process, the 350 analyzed instances were used to obtain the new formula for the “Sousa Valley Best Ear” competition. This last step was done for the best interpretable

regression method according to the PR.NDCG evaluation measure developed in the second step.

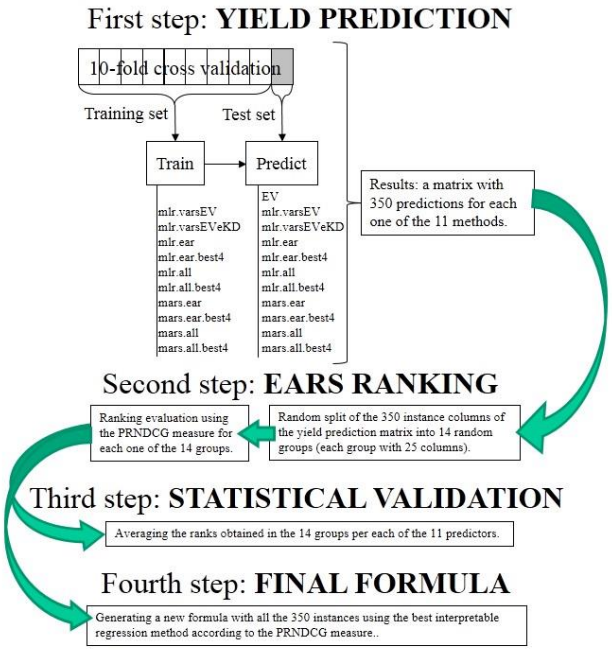

Figure VI.1 Steps of the experimental procedure: (1) Yield prediction, 350 instances were divided into ten subsets from which 9 sets were used for training and 1 set for prediction. This process was repeated 10 times generating predictions for the 350 instances using each of the 11 methods; (2) ears ranking, each matrix composed of 350 instances were divided into 14 disjoint matrix (each one with 25 instances) per each of the eleven methods allowing the evaluation ranking using PRNDCG; (3) statistical validation of average differences between the PRNDCG results using the Friedman rank test and multiple comparisons using the Nemenyi post hoc tests; (4) final formula, after method selection we have used the initial 350 instances to determine the formula.

Subsequently, the data from 13 previous editions of the "Sousa Valley Best Ear" competition were used to compare the ranks obtained by the EV formula and the new developed formula.

This section is organized according to the four steps above described.

VI.6.3.1 Yield prediction

Since the goal of yield prediction is to predict the yield for a given new ear, the prediction ability should be tested using ears that were not used for training the model. With this purpose, we have used as resampling method, 10-fold cross-validation (Stone 1974) on the 350 instances. This method randomly splits the given instances into 10 subsets of equal size (35 instances in each). Then, it trains the model using the instances from 9 subsets, and tests it in the remainder subset. This process is repeated 10 times always leaving a different subset for testing. Using our data, it means 315 instances for training and 35 for testing at each of the 10 iterations.

This procedure assures that the quality of the model is evaluated according to its prediction ability on the test set. The model is created even if the assumptions are not met, namely the linearity or the normality assumptions. When these assumptions are not guaranteed, the prediction evaluation on the test set is expected to be poorer.

Two different algorithms were used to predict yield: MLR and MARS (as described in Section VI.6.2.1). Each algorithm was tested twice, using all 32 variables (24 ear's and 8 plant's field trial variables), or just the ear variables, plus yield. Naturally, for the ear competition, field variables cannot be used. However, its introduction in the analysis aimed to obtain additional information about the relation between these field variables and yield, what could be useful for

future farmers' selection. The traits selection was done by forward search using as stop criterion the increase of the squared error in the test set. The evaluation of the error in the test set is expected to overestimate the accuracy. Subsequently, MLR and MARS combinations were also tested allowing a maximum of 4 variables/terms in order to study the possibility of obtaining acceptable results with a limited number of traits. We have used four variables (with MLR) or terms (with MARS) because this is also the number of variables used in the EV formula. This gave raised to 8 variants of interpretable methods to be tested. With the MLR algorithm, two more variations were tested. One of them uses the variables in the EV formula (mlr.varsEV). The other (mlr.varsEVeKD) uses the same variables as in the EV formula plus the KD (kernel dept) variable that was previously described as an important variable for yield prediction in literature (Hallauer and Carena 2009). The different algorithms with all 11 described variations are presented in Table VII.6.

All experiments were done using the R-project (Team 2011). We have used the lm function with the default parameters (Team 2011) as MLR implementation. For the MARS implementation we used the earth function available in the earth package (Milborrow 2011). This function has two important parameters: nk, which is the maximum number of terms that are allowed in the MARS formulation; and degree, that defines the maximum number of interactions. These parameters are explained in Section VI.6.2.3. We have used nk = 120

as maximum number of terms, and degree = 3 as maximum number of interactions. These values were chosen as the best ones after initial experimentations.

Table VI.6Different interpretable methods for yield prediction

Nr.	Name	Algorithm	Initial feature set	Search	Max limit
1	EV	EV formula	Fixed set={KW,L,R,KN}	*n.a.*	*n.a.*
2	mlr.varsEV	MLR	Fixed set={KW,L,R,KN}	*n.a.*	*n.a*
3	mlr.varsEVeKD	MLR	Fixed set={KW,L,R,KN,KD}	*n.a.*	*n.a.*
4	mlr.ear	MLR	Ear variables	Forward	*n.a.*
5	mlr.ear.best4	MLR	Ear variables	Forward	4
6	mlr.all	MLR	Ear + field variables	Forward	*n.a.*
7	mlr.all.best4	MLR	Ear + field variables	Forward	4
8	mars.ear	MARS	Ear variables	Forward	*n.a.*
9	mars.ear.best4	MARS	Ear variables	Forward	4
10	mars.all	MARS	Ear + field variables	Forward	*n.a.*
11	mars.all.best4	MARS	Ear + field variables	Forward	4

[n.a.] means not applicable

We have used the coefficient of variation as evaluation measure for the yield predictions. Its values are expressed as the ratio between the squared root of the mean squared error (mse) and the yield average.

$\text{varIndex} = \sqrt{mse} / \bar{y}$, where $\bar{y}$ is the yield average

Equation VI.16

$$mse = \frac{1}{n}\sum_{i=1}^{n}(py_i - y_i)^2$$

Equation VI.17

where py_i is the yield prediction and y_i is the true yield value for the instance i.

where py_i is the yield prediction and yi is the true yield value for the instance i.

Consequently, it is easily interpretable because its value is given as a percentage of the yield average.

VI.6.3.2 Ears ranking

In a second step, the ears were ranked according to the predicted yield using the different algorithms variations (as presented in Table VII.6). For that, the matrix with 350 yield predictions columns for each one of the 11 methods, were divided into 14 random groups (each with 25 elements). By choosing 14 groups it is possible to guarantee simultaneously an adequate amount of groups for the statistical validation (to be discussed in Section VI.6.3.3) and enough number of instances per group, 25. The use of 25 instances per group is sufficient considering that only the 10 first ranked instances according to both ranks, the observed yield rank and the predicted yield rank, are considered for evaluation. For each group and for each method, the instances were ranked using several instance ranking evaluation measures previously discussed (Section VI.6.2.4). Despite

we have developed and chosen as the most appropriated approach for evaluating the different ranks the PR.NDCG measure (Equation VI.3), the results obtained using the variation index (Equation VI.16), the R.NDCG (Veldboom, Lee 1996a) (Equation VI.15) and the P.NDCG (Veldboom, Lee 1996a) (Equation VI.2) measures are also presented in order to better discuss the PR.NDCG ranking results.

VI.6.3.3 Statistical validation

We have compared the eleven algorithms variations developed using the PR.NDCG instance ranking evaluation measure. The statistical validation of these results was done using the Friedman rank test with the statistic derived by Iman and Davenport as described by (Demsar 2006). The assumed level of significance considered for the null hypothesis of equivalence between the eleven ranking methods for p-value was 0.01. The post hoc Nemenyi test was used to validate whether the difference of the averaged ranks of any two ranking methods is larger enough to be statistically valid for the desired level of significance.

VI.6.3.4 Ear value formula generation and "Sousa Valley Best Ear" competition

The new formula to be used for the "Sousa Valley Best Ear" competition was obtained using the top ranked algorithm and variation in the PR.NDCG rank, as discussed in Section VI.6.3.2. The new formula used all the 350 analyzed instances. Theoretically, this is a more stable model in opposition to the 10 different formulas

obtained by the cross validation process. In this way, this new formula is a better estimate of the yield potential using ear traits for selection and also the best formula for the ear competition.

Subsequently, the data from 13 previous editions of the "Sousa Valley Best Ear" competition were used to compare the ranks obtained by the original EV formula and the new developed formula. In this comparison it was accounted how many top five to ten ranking positions from the original EV were maintained on the top five to ten ranking positions of the newly developed formula. In addition and as an example, the range values of the newly developed EV formula were calculated for standard maize germplasm, so as the associated value traits.

VI.7. Acknowledgements

Conception: SP, ARH

Design of the work: JMM, PMM, MCVP

Acquisition of data: PMM, AF, EA, SP

Analysis and interpretation of data: PMM, JMM

Article drafting: PMM, MCVP, JMM

Revising it critically: ARH, SP, MCVP, JMM

Populations' development and breeding: SP, farmer FM, farmers with their germplasm at "Sousa Valley best ear competition"

VI.8. References

Badstue LB, Bellon MR, Berthaud J, Ramírez A, Flores D, Juárez X (2007) The dynamics of farmers' maize seed supply practices in the central valleys of Oaxaca, Mexico. World Dev. 35: 1579–1593. doi:10.1016/j.worlddev.2006.05.023

Baker RJ (1986) Selection Indices in Plant Breeding. CRC Press, Inc. Bowman, M., Crossley, B., 1908. The varieties of dent corn., pp. 424–446, Corn growing.

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and Regression Tree. Chapman and Hall/CRC.

Busch W, Benfey PN (2010) Information processing without brains – the power of intercellular regulators in plants. Development 137: 1215-1226. doi: 10.1242/dev.034868

Dawson JC, Serpolay E, Giuliano S, Schermann N, Galic N, Berthellot JF, Chesneau V, Ferté H, Mercier F, Osman A, Pino S, Goldringer I (2013) Phenotypic diversity and evolution of farmer varieties of bread wheat on organic farms in Europe. Genetic Resources and Crop Evolution 60: 145-163. doi: 10.1007/s10722-012-9822-x

Demsar J (2006) Statistical comparisons of classifiers over multiple data sets. J. Mach. Learn. Res. 7: 1–30.

Doebley J (2004) The genetics of maize evolution. Annu. Rev. Genet. 38: 37–59. doi: 10.1146/annurev.genet.38.072902.092425

Fitzgerald D (1993) Farmers deskilled: hybrid corn and farmers' work. Technol. Cult. 34: 324–343.

Friedman JH (1991) Multivariate adaptive regression splines. Ann. Stat. 19: 1–141.

Fürnkranz J, Hüllermeier E (2011) Preference Learning and Ranking by Pairwise Comparison. Preference learning. Springer, pp. 65–82.

Hallauer AR, Carena MJ (2009) Maize breeding. In: Carena, M.J. (Ed.), Handbook of Plant Breeding, Cereals. Springer, New York, pp. xiv, 425.

Hallauer AR, Carena MJ, Miranda Filho JB (2010a) Quantitative genetics in maize breeding. Springer, New York. doi:10.1007/978-1-4419-0766-0

Hallauer AR, Ross AJ, Lee M (2010b) Long-Term Divergent Selection for Ear Length in Maize. John Wiley & Sons, Inc.

Järvelin J, Kekäläinen J (2000) IR evaluation methods for retrieving highly relevant documents. In: ACM SIGIR, pp. 41–48.

Järvelin J, Kekäläinen J (2002) Cumulated gain-based evaluation of IR techniques. ACM Trans. Inf. Syst. 20: 422–446. doi: 10.1145/582415.582418

Jones DF (1935) The similarity between fasciations in plants and tumors in animals and their genetic basis. Science 81: 75–76

Kazai G, Lalmas M (2006) eXtended Cumulated gain measures for the evaluation of content-oriented XML retrieval. ACM Trans. Inf. Syst. 24: 503–542. doi: 10.1145/1185877.1185883

Kleinbaum DG, Nizam KLL, Muller AKE (2008) Applied Regression Analysis and Other Multivariable Methods. Thomson.

Klesselbach T (1922) Ear-type selection and yield of dent corn. Agron. J. 14: 27–48

Lin CY (1978) Index selection for genetic improvement of quantitative characters. Theor. Appl. Genet. 52: 49–56. doi: 10.1007/BF00281316

Louette D, Smale M (2000) Farmers' seed selection practices and traditional maize varieties in Cuzalapa, Mexico. Euphytica 113: 25–41. doi: 10.1023/A:1003941615886

Machado AT, Nass LL, Machado TTC (2011) Manejo sustentavel agrobiodiversidade nos biomas Cerrado e Caatinga. EMBRAPA.

Manning CD, Raghavan P, Schütze H (2008) Introduction to Information Retrieval. Cambridge University Press, Cambridge.

Mendes Moreira PMR, Pêgo SE, Vaz Patto MC, Hallauer AR (2008) Comparison of selection methods on 'Pigarro', a Portuguese improved maize population with fasciation expression. Euphytica 163: 481-499. doi: 10.1007/s10681-008-9683-8

Mendes Moreira PMR, Pêgo SE, Vaz Patto MC, Hallauer AR (2008) Comparison of selection methods on 'Pigarro', a Portuguese improved maize population with fasciation expression. Euphytica 163: 481-499. doi: 10.1007/s10681-008-9683-8

Milborrow S (2011) Derived from mda:mars by Trevor Hastie SMD and Tibshirani. R. Earth: Multivariate Adaptive Regression Spline Models.

Moreira PM (2006) Participatory maize breeding in Portugal. A case study. Acta Agronomica Hungarica 54: 431–439. doi: http://dx.doi.org/10.1556/AAgr.54.2006.4.6

Pêgo SE, Hallauer AR (1984) Portuguese maize germplasm with abnormal ear shape. Maydica 29: 39–53

Piwowarski B, Dupret G (2006) Evaluation in (XML) information retrieval: expected precision-recall with user modelling (EPRUM). In: 4th International Workshop of the Initiative for the Evaluation of XML Retrieval, pp. 30–42.

"Prasad A, Iverson L, Liaw A (2006) Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems 9: 181–199. doi:

10.1007/s10021-005-0054-1"

Pressoir G, Berthaud J (2004) Population structure and strong divergent selection shape phenotypic diversification in maize landraces. Heredity 92: 95–101. doi:10.1038/sj.hdy.6800388

Shull GH (1908) The Composition of a Field of Maize., pp. 296–301.

Shull GH (1909) A Pure-Line Method of Corn Breeding., pp. 51–59.

Smith HF (1936) A discriminant function for plant selection. Ann. Eugen. 7: 240–250

Soleri D, Cleveland DA (2009) Breeding for quantitative variables. Part 1: Farmers' and scientists' knowledge and practice in variety choice and plant selection. Plant breeding and farmer participation. Food and Agriculture Organisation of the United Nations, 323–366

Soleri D, Smith SE, Cleveland DA (2000) Evaluating the potential for farmer and plant breeder collaboration: a case study of farmer maize selection in Oaxaca, Mexico. Euphytica 116: 41–57

Stone M (1974) Cross-validatory choice and assessment of statistical predictions. J. R. Stat. Soc. Ser. B 36: 111–147

Taguchi-Shiobara F, Yuan Z, Hake S, Jackson D (2001) The *fasciated ear2* gene encodes a leucine-rich repeat receptor-like protein that regulates shoot meristem proliferation in maize. Genes Dev 15: 2755-2821. PMID: 11641280

Team, RDC (2011) R: A Language and Environment for Statistical Computing.

"Turpin KM, Lapen DR, Gregorich EG, Topp GC, Edwards M, McLaughlin NB, Curnoe WE, Robin MJL (2005) Using multivariate adaptive regression splines

(MARS) to identify relationships between soil and corn (*Zea mays* L.) production

properties. Can. J. Soil Sci. 85: 625–636. doi: 10.4141/S04-062"

Vaz Patto MC, Moreira PM, Carvalho V, Pêgo S (2007) Collecting maize (*Zea mays* L. convar. *mays*) with potential technological ability for bread making in Portugal. Genet Res Crop Evol 54:1555‑1563. doi: 10.1007/s10722-006-9168-3

Villordon A, Clark C, Ferrin D, LaBonte D (2009) Using growing degree days, agrometeorological variables, linear regression, and data mining methods to help improve prediction of sweetpotato harvest date in Louisiana. HortTechnology 19: 133–144

Wilkes G (2004) Corn, strange and marvelous: but is a definitive origin known? In: Wayne Smith, C. (Ed.), Corn: Origin, History, Technology, and Production. Wiley & Sons, Inc., pp. 3–63.

Williams J (1962) The evaluation of a selection index. Biometrics 18: 375–393. doi: 10.2307/2527479

Winter F (1925) The effectiveness of seed corn selection based on ear characters. Agron. J. 17: 113–118

CHAPTER VII.

Genetic Architecture of Ear Fasciation in Maize (*Zea mays*) under QTL Scrutiny

The work presented in this chapter was mostly performed by Pedro Mendes Moreira (see acknowledgments section) and partially included in the following publication: Mendes-Moreira P, Alves ML, Šatović Z, dos Santos JP, Santos JN, Souza JC, Pêgo SE, Hallauer AR, Vaz Patto MC (2015) Genetic Architecture of Ear Fasciation in Maize (*Zea mays*) under QTL Scrutiny. PLoS ONE 10(4): e0124543. doi:10.1371/journal.pone.0124543.

VII.1. Abstract

Fasciation expresses phenotypically as an uncontrolled meristematic activity, parallel to cancer development in animals. This proliferation activity, often considered as a negative trait, can have its pay off. For this reason the knowledge of the genes affecting maize ear fasciation may lead to better grain yield modeling.

The importance of fasciation has been a topic of interest via scientific journals and patents. Great variability within ear fasciation exists in the Portuguese material because Portuguese farmers considered fasciation as an interesting component for yield improvement and for this reason they have kept on selecting it along generations. National maize breeders were influenced by this farmer intuition and included fasciation in their inbred lines development programs.

The present study pursued the following aims: (1) To determine the genetic relationships between a comprehensive set of ear architecture traits related with fasciation in a segregating F2 population, developed from a cross between contrasting (non-fasciated PB260 x fasciated PB266) inbred lines selected in Portugal, (2) to identify chromosomal positions, size and effects of QTLs (Quantitative Trait *Loci*) involved in the inheritance of those traits, across two environments, using univariate and multivariate approaches and (3) to identify possible candidate genes associated with these QTL.

We have detected significant variation for maize ear fasciation and related ear traits and mapped a number of QTLs controlling those

traits in the Portuguese derived PB260 x PB266 segregating population. This work revealed some already know chromosomal ear fasciation control related regions, where some candidate genes are already identified, but also unravel new regions where new candidate genes can exist.

This is a very interesting trait for maize breeding worldwide, but one that must be fully understood at a genetic level before perfectly controlled in breeding programs. This control can be attained by the development of molecular selection tools based on QTL flanking molecular markers or associated functional markers (candidate genes), such as the ones identified in the present study.

VII.2. Introduction

"Fasciation" derives from the Latin word fascis, meaning "bundle" and is a reflection of increased cell proliferation (Bommert et al. 2005). One of the earliest reports of fasciation in maize dates from 1912 and, at that time, fasciated maize ears were frequently found in US maize fields (Emerson 1912), mostly in dent and pop maize germplasm (White 1948). There is a widespread occurrence of fasciated variants among vascular plants, and it has been reported that it increases crop yields (White 1948). Meristematic activity in the inflorescence has a profound influence on grain yield. Grain yield in maize is a complex, continuous trait that might be modified by a large number of genes including those controlling ear architecture traits. A

thorough knowledge of the genes affecting these various components would lead to better yield modeling.

Many forward genetic screens have uncovered strong loss-of-function mutants with several altered maize ear architecture types and their responsible genes (Pautler et al. 2013). Examples are the *fasciated2* (*fea2*) (Taguchi-Shiobara et al. 2001; Bommert et al. 2013), *compact plant2* (*ct2*) (Eckardt 2007), *double cob1* (*dbcb1*) (Brewbaker 2009) and the ramosa genes (*ra1*, *ra2*, and *ra3*) (Neuffer et al. 1997).

While these mutants are useful for uncovering the normal function of genes, they rarely provide useful material for breeding efforts because they frequently display negative pleiotropic traits (Pautler et al. 2013). For example, the *fea2*, characterized by an increased kernel row number, is associated with a decrease in the length of the ear, as well as disorganized seed rows that limit the number of seeds per ear and their harvesting ability (Pautler et al. 2013). Additionally, much of the natural variation in inflorescence shape observed in maize is due to the cumulative effect of several *loci*. The responsible genes for the milder quantitative variation in these ear architecture traits or weaker alleles of these strong mutant variants will be particularly suitable for direct use on yield breeding approaches through Marker Assisted Selection (MAS). Many maize QTL studies have focused on ear architecture traits, with special interest on kernel row number (Veldboom, Lee 1996a; Upadyayula et al. 2006; Lu Y et al. 2009; Zhang et al. 2010; Brown et al. 2011; Jiao Y et al. 2012) but just a few have attained fine mapping of major detected QTL (Bommert et al.

2013; Liu et al. 2012; Steinhoff et al. 2012; Zhang et al. 2013). Bommert et al. (2013) have in fact isolated a weak allele of the *fea2* and showed that this allele increases kernel row number and number of kernels per ear, without causing a fasciated or shorter ear.

Since maize introduction to the country in the 15th century, after Columbus, the importance of maize ear fasciation was quickly understood by Portuguese farmers who saw it as a way to improve production (Ferrão 1992). In traditional Portuguese maize landraces, ears are often found abnormally flattened and wider than normal, sometimes with irregular seed rows, but not particularly short in length. In addition to robustness and yield stability, Portuguese farmers preferred to select for large size ears without regard to shape, maintaining a certain level of diversity. This ear trait phenotype, known as "bear's foot" in English, (Kempton 1923), corresponds to several popular names in Portuguese ("pé-de-porco", "pata de porco","unha-de-porco", "mão de morto", "milho espalmado", "mãozeira" or "milho das mãozinhas"), highlighting the importance of this trait for Portuguese farmers.

Contrary to other domestication and crop improvement traits (Yamasaki et al. 2005), diversity in this ear trait was preferred and maintained by Portuguese farmers as an important parameter influencing yield (Vaz Patto et al. 2003). In fact, fasciation trait expression varies with the environment, *i.e.*, more inputs induce higher fasciation expression (for example, lower plant densities and more available nutrients) (White 1948; Mendes-Moreira et al. 2014).

During the Portuguese maize collecting expedition in 2005 (Vaz Patto et al. 2003), 56% of the traditional maize landraces collected had some degree of fasciation, *versus* the 10% observed during the previous collecting missions in the 1980s. This fact indicates farmers' preferences related to adaptation of their traditional agricultural systems: *i.e.*, they selected a germplasm with enough yield plasticity and wider adaptability to different crop systems. Because of this finding, we might consider Portuguese maize landraces as a diversifying germplasm extremely important to seek novel alleles for breeding. Indeed the ear fasciation trait has already been exploited in a maize Portuguese Participatory Plant Breeding (PPB) program where as much as 18 up to 22 kernel rows per ear were obtained in the improved Open Pollinated Varieties (OPV) (*e.g.* 'Pigarro' or 'Fandango').

Early genetic studies of six Portuguese maize traditional landraces with a high frequency of abnormal ear shape (high level of fasciation expression) crossed with ramosa mutants indicated that fasciation was not associated with the ramosa genes (*ra1*, *ra2* or *ra3*) and a complex system of inheritance was proposed (Pêgo 1982). Additionally, this typical Portuguese ear fasciation was considered as a useful trait for improving yield when intermediate expression was attained, in order to allow certain uniformity in the ear and the plant (Pêgo, Hallauer 1984). Unfortunately, the genetic control of this ear trait was not elucidated, decreasing the possibilities of using it in an efficient and fast maize breeding approach.

The development of molecular markers allows us to study the genetic basis of complex quantitative traits, such as the typical Portuguese ear fasciation, in further detail and to develop tools for sustaining modern breeding approaches.

In the present study we: (1) determine the genetic relationships between a comprehensive set of ear architecture traits related with fasciation in a segregating $F_{2:3}$ population, developed from a cross between contrasting (non-fasciated PB260 x fasciated PB266) inbred lines selected in Portugal, (2) identify chromosomal positions, size and effects of QTL involved in the inheritance of those traits, across two environments, using univariate and multivariate approaches and (3) identify possible candidate genes associated with these QTL.

VII.3. Results

VII.3.1 Genetic variation, heritabilities and phenotypic correlations

The parental accession PB266 had an average phenotypic value, for ear diameter 3 and 4, row number 2, cob/ear weight per ear and cob diameter 4, significantly higher than the parental accession PB260. Additionally, PB266 was also significantly more fasciated than PB260 (respectively 2.38 *versus* 1.41, near double) (Table VII.1, Table VII.2).

Genotypic effects were highly significant for all investigated traits ($P < 0.01$). Nevertheless, genotype x environment interaction was significant for the majority of the traits, with the exception of ear length, fasciation, convulsion, cob/ear weight per ear, kernels per

row, ear diameter 2, medulla 1 and 2 and rachis 2 (Table VII.1, Table VII.2). Because of these findings, and to better understand the stability of the QTL across multiple environments, QTL mapping was performed separately for each environment data set.

$F_{2:3}$ families showed a large and continuous variation for the investigated traits. The Kolmogorov- Smirnov's test of normality showed that a few of the studied traits were not normally distributed, such as the fasciation trait (Table VII.1, Table VII.2). Transformation of fasciation data by the natural logarithm function improved normality, but had little effect on the analyses. Hence results were described only for the untransformed data.

In general, the phenotypic values of the two parental inbred lines were significantly different from at least one of the most extreme F2 plants ($P \leq 0.05$), and transgressive segregation was observed for all the traits with the exception of kernel depth and ear length. However, these comparisons should be considered with caution, since the presented data from parental lines and $F_{2:3}$ families were obtained in different environmental conditions (Table VII.2), during several years of evaluation, at least for the parental accessions.

Table VII.1 Traits measured, codes and respective description of measurements.

Trait #	Codes	Name	Units/ Scale	Data/ plot	Measurement description
1	yld	Yield	Mg ha-1	1	Grain yield 15% moisture (Mg ha^{-1}) = total Ear weight per ha x cwew x (100% - % moisture at harvest)/(100%-15%moisture) Grain moisture measured with the FARMPOINT® moisture meter using amixed sample of four shelled ears grain.
2	cwew	Cob/ear weight at harvest	%	1	Ratio of cob weight in the ear weight per plot (sample of 4 ears)
3	en	Ears number	nº	1	Number of ears per square meter
4	av_ew	Ear weight at harvest	g	1	Average ear weight corrected for 15% moisture
5	l	Ear length	cm	5	Ear length
6-7	ed1, ed3	Ear diameter 1 and 3	cm	5	Large diameter in the 1/3 bottom and top of the ear respectively
8-9	ed2, ed4	Ear diameter 2 and 4	cm	5	Small diameter in the 1/3 bottom and top of the ear respectively (90º rotation from large diameter)
10-11	cd1, cd3	Cob diameter 1 and 3	cm	5	cd1 and 3 measure in the same way for ed's
12-13	cd2, cd4	Cob diameter 2 and 4	cm	5	cd2 and 4 measure in the same way for ed's
14	kd	Kernel dept	cm	5	Kernel dept, one kernel in the middle of the ear
15-16	m1, m2	Medulla 1 and 2	cm	5	Large and small length of medulla, respectively ,cob is cut in the Diameter 1 position [78,79]
17-18	rq1, rq2	Rachis 1 and 2	cm	5	Large and small length of rachis; cob is cut in the Diameter 1 position [78,79]
19	Ew	Ear weight	g	5	Ear weight, adjusted to 15% of grain moisture
20	Cw	Cob weight	g	5	Cob weight, adjusted to 15% grain moisture
21	Sw	Thousand-kernel weight	g	5	Thousand-kernel weight at 15% grain moisture
22	Kw	Kernel weight	g	5	kernel weight per ear, adjusted to 15% grain moisture
23	e_cwew	Cob/ear weight per ear	%	5	Percentage of cob weight in the ear weight measured per ear at lab
24-25	r1, r2	Kernel-row number 1 and 2	nº	5	Row number in the 1/3 bottom and top of the ear respectively
26	Fa	Fasciation	1 to 9	5	Fasciation degree (1 – without fasciation and 9 as a maximum of fasciation)
27	Cv	Ear convulsion	0 to 5	5	Convulsion intensity, kernel-row arrangement in the ear (0 - without convulsion, regular kernel-row arrangement, 5 – maximum of convulsion, without kernel-row arrangement)
28	Kn	Kernel number	nº	5	Kernel number per ear
29	Kr	Kernel per row	nº	5	Kernel number per row

Table VII.2 Phenotypic values (mean ± standard deviation) of the parental inbred lines PB260 and PB266, respective $F_{2:3}$ families and quantitative genetic parameters for ear fasciation and related traits

Trait	PB260[a]	PB266		$F_{2:3}$ PB260 x PB266						ANOVA[f]			
	c[b]	c	Sheffe[c]	c	m	c&m	c	m	c&m				
	χ ± sd	χ ± sd		χ ± sd; KS[d]	χ ± sd; KS	r[e]	h^2	h^2	h^2	G	E	REP(E)	G x E
yld	0.53	0.41		3.9±1.24	4.04±1	0.44***	0.67	0.48	0.61	***	ns	ns	*
cwew				0.26±0.06**	0.29±0.04*	0.22*	0.54	0.31	0.24	***	ns	***	**
en				4.44±0.98**	4.62±0.84**	0.35***	0.66	0.37	0.51	***	ns	ns	**
av_ew				82.0±16.0	87.0±19.5	0.29**	0.50	0.42	0.42	***	ns	*	*
l	11.24±1.57	9.24±2.11	$PB266^{a}<minF_2^{ab}<PB260^{ab}<MaxF_2^{b}$	13.16±1.67	14.14±1.6	0.71***	0.79	0.73	0.72	***	*	ns	ns
ed1	3.59±0.21	3.87±0.64	$PB260^{a}<minF_2^{a}<PB266^{a}<MaxF_2^{b}$	4.72±0.32	4.81±0.31	0.67***	0.76	0.71	0.78	***	ns	***	*
ed3	3.1±0.29	3.88±0.76	$PB260^{a}<minF_2^{a}<PB266^{b}<MaxF_2^{c}$	4.43±0.32	4.38±0.35	0.63***	0.81	0.63	0.76	***	ns	**	*
ed2	3.56±0.39	3.44±0.55	$minF_2^{a}<PB266^{a}<PB260^{a}<MaxF_2^{b}$	4.47±0.33	4.58±0.3	0.65***	0.75	0.68	0.75	***	ns	***	*
ed4	3.04±0.37	3.58±0.73	$PB260^{a}<minF_2^{ab}<PB266^{b}<MaxF_2^{c}$	4.26±0.3	4.2±0.31	0.54***	0.79	0.53	0.69	***	ns	**	*
r1	10.24±2.54	10.94±3.87	$PB260^{a}<PB266^{a}<minF_2^{b}<MaxF_2^{c}$	16.6±1.57	16.83±1.73	0.7***	0.76	0.77	0.81	***	ns	ns	*
r2	9.65±2	13.5±4.94	$PB260^{a}< minF_2^{ab}<PB266^{b}< MaxF_2^{c}$	16.67±1.6	16.72±1.85	0.72***	0.79	0.78	0.83	***	ns	ns	*
fa	1.41±0.51	2.38±0.62	$minF_2^{a}<PB260^{b}<PB266^{c}<MaxF_2^{d}$	2.41±0.67**	2.09±0.6**	0.67***	0.61	0.70	0.73	***	ns	***	ns
cv	2.59±0.71	2.44±1.09	$minF_2^{a}<PB266^{b}<PB260^{b}<MaxF_2^{c}$	2.31±0.4*	2.21±0.36**	0.44***	0.49	0.35	0.59	***	ns	***	ns
kd	0.73±0.1	0.88±0.1	$PB260^{a}<PB266^{ab}<minF_2^{ab}<MaxF_2^{b}$	1.02±0.06*	1±0.05	0.64***	0.84	0.65	0.75	***	*	ns	**
ew	47.54±16.23	46.12±24.11	$PB266^{a}<PB260^{a}<minF_2^{a}<MaxF_2^{b}$	127.04±25.12	133.07±23.21	0.53***	0.72	0.56	0.68	***	ns	**	*
kw	37.06±14.87	33.02±19.18	$PB266^{a}<PB260^{a}<minF_2^{a}<MaxF_2^{b}$	100.53±19.36	102.32±17.24	0.52***	0.71	0.51	0.68	***	ns	**	*
cw	10.42±3.61	12.89±6.86	$minF_2^{a}<PB260^{a}<PB266^{a}<MaxF_2^{b}$	26.51±7.15	30.75±7.21	0.64***	0.75	0.73	0.68	***	*	ns	*

Trait	PB260[a]	PB266	Sheffe[c]	$F_{2:3}$ PB260 x PB266									
	c[b]	c		c	m	c&m	c	m	c&m	ANOVA[f]			
	χ ± sd	χ ± sd		χ ± sd; KS[d]	χ ± sd; KS	r[e]	h^2	h^2	h^2	G	E	REP(E)	G x E
e_cwew	0.23±0.09	0.31±0.15	$minF_2^a$<PB260[b]<PB266[c]<$MaxF_2^c$	0.21±0.03	0.23±0.03	0.77***	0.73	0.82	0.69	***	*	ns	ns
kn	151.51±54.36	139.68±35.65	PB266[a]<PB260[a]<$minF_2^a$<$MaxF_2^b$	353.61±68.86	389.68±60.12	0.59***	0.76	0.52	0.64	***	ns	**	ns
sw	244.3±43.78	226.74±100.4	$minF_2^a$<PB266[b]<PB260[b]<$MaxF_2^c$	288.06±30.72	264.04±28.48	0.64***	0.81	0.67	0.58	***	***	ns	*
kr	15.18±3.81	12.38±3.61	PB266[a]<$minF_2^a$<PB260[a]<$MaxF_2^b$	23.36±3.49	25.76±3.02	0.68***	0.78	0.67	0.64	***	*	ns	ns
cd1	2.72±0.36	2.7±0.47	$minF_2^a$<PB266[a]<PB260[a]<$MaxF_2^b$	3.2±0.26	3.33±0.26	0.73***	0.72	0.79	0.77	***	ns	**	ns
cd3	2.1±0.19	2.51±0.5	$minF_2^a$<PB260[a]<PB266[a]<$MaxF_2^b$	2.82±0.23	2.77±0.27	0.68***	0.80	0.71	0.79	***	ns	*	*
cd2	2.56±0.4	2.4±0.44	$minF_2^a$<PB266[ab]<PB260[b]<$MaxF_2^c$	2.91±0.23	3.04±0.21*	0.68***	0.69	0.71	0.71	***	ns	***	ns
cd4	2.01±0.21	2.33±0.44	$minF_2^a$<PB260[a]<PB266[b]<$MaxF_2^c$	2.62±0.19	2.59±0.21	0.53***	0.74	0.57	0.69	***	ns	*	*
m1	0.73±0.2	0.66±0.36	PB266[a]<PB260[a]<$minF_2^a$<$MaxF_2^b$	1.32±0.17	1.39±0.21	0.67***	0.64	0.76	0.76	***	ns	*	ns
m2	0.62±0.19	0.5±0.26	PB266[a]<$minF_2^a$<PB260[a]<$MaxF_2^b$	1.02±0.14	1.06±0.16	0.67***	0.69	0.60	0.78	***	ns	*	ns
rq1	1.89±0.18	1.8±0.32	$minF_2^a$<PB266[a]<PB260[a]<$MaxF_2^b$	2.32±0.23	2.46±0.23	0.67***	0.72	0.79	0.68	***	ns	***	**
rq2	1.62±0.2	1.54±0.38	$minF_2^a$<PB266[ab]<PB260[b]<$MaxF_2^c$	1.97±0.21	2.11±0.2	0.66***	0.72	0.68	0.67	***	ns	***	ns

[a] PB260 and PB266 data obtained from 2010 and 2012 organic production field trial, with 2 replications at Coimbra. $F_{2:3}$ data obtained from two conventional field trials at Coimbra and Montemor, with two replications.

[b] c - Coimbra and m - Montemor

[c] $minF_2$– top five minimum values of F_2 PB260 x PB266; $MaxF_2$– top five maximum values of $F_{2:3}$ PB260 x PB266. Significant differences exist when no letter repetition occurred between groups

[d] KS - *, ** - Significance of the Kolmogorov-Smirnov's test of normality

[e] r - correlation between the two environments per trait. h^2 – broad sense heritabilities

[f] Significance of the sources of variability: G - Genotype, E - Environment, Rep (E) - Repetitions within Environment, G x E - Genotype x Environment Interaction

Levels of significance: ns non-significant value; * significant at P < 0.05; ** significant at P < 0.01; *** significant at P < 0.001

Traits measured: yld –yield; cwew - cob/ear weight at harvest; en - ears number; av_ew – 20 ears average weight at harvest t; l – ear length; ed 1 to 4 - ear diameter 1 to 4; cd1 to 4 - cob diameter 1 to 3; kd - kernel dept; m1, m2 - medulla 1 and 2; rq1, rq2 -rachis 1 and 2; ew -ear weight; cw -cob weight; sw -thousand kernel weight; kw - kernel weight; e_cwew - cob/ear weight per ear; r1, r2 - kernel-row number 1 and 2; fa - fasciation; cv - ear convulsion; kn - kernel number; kr - kernel per row.

Broad-sense heritabilities of investigated traits across both environments were in general above or equal to 0.70, apart from grain yield, the cob/ear weight at harvest, number of ears, average ear weight at harvest, ear convulsion, kernel number, thousand kernel weight and number of kernel rows. In general, broad sense heritabilities were higher at the Coimbra than Montemor environment with the exception of rachis 1, medulla 1, cob diameters 1 and 2, cob/ ear weight per ear, kernel row numbers and fasciation. Fasciation heritability was 0.73 across both environments and ranged from 0.61 at Coimbra to 0.70 at Montemor (Table VII.2).

Trait correlations between the two environments were highly significant ($P < 0.001$) with the exception of cob weight ($P < 0.01$) and rachis 2 ($P < 0.05$) (Table VII.2). Correlation coefficients range from 0.22 to 0.77 (rachis 2 and thousand kernel weight, respectively), but were above 0.60 for many traits, as in the case of fasciation (0.67) and ear diameter 3 (0.63). High correlations (0.70 to 1.00) were observed for ear length (0.71), row number 1 and 2 (0.70 and 0.71), cob/ear weight per ear (0.77) and cob diameter 1 (0.73) (Table VII.2).

The highest correlations (0.9–1.00) observed in both environments were among ear diameters from the base of the ear or from the top of the ear (ear diameters 1 and 2 and ear diameters 3 and 4, respectively), rows number 1 and 2 and cob diameter 1 and 2 with rachis 1 and 2, respectively. The rachis and medulla had high correlations, as was expected, because their differences represent

the thickness obtained by glumes, paleas and lemmas components. In addition, ear and kernel weight were also highly correlated.

The fasciation trait at Coimbra and Montemor was correlated with ear diameter 3 (0.53 and 0.76), row number 2 (0.45 and 0.75) and cob diameter 3 (0.59 and 0.79). These two last traits presented the highest correlations between ear diameters and row numbers. Additionally, correlations varying between 0.70–0.90 were detected for both environments for medulla 1 and 2 and rachis 1 and 2, and also among and within ear diameters (ear diameters 1 and 2 with ears diameters 3 and 4) and cob diameters (cob diameters 1 with 2 and 3, and cob diameter 3 with 4). High correlations among rachis 1 and cob diameter 2 and medulla 1 were also observed. The yield and cob weight were correlated with ear and kernel weight. Finally, kernel number was correlated with ear and kernel weight and number of kernels per row. This last trait was also correlated with ear length.

The three principal components explained 73.95% and 71.1% of the variation, respectively, at the Coimbra and Montemor environments. Principal component 1 (PC1) for Coimbra explained 43.61% of the total variation present in the data set. For this PC1, the traits that contribute the most for explaining variation were ear diameter 1 and 2 (correlation, 0.90 to 1.00) and ear diameter 3 and 4, ear, kernel and cob weight, cob diameters, medulla and rachis correlation, 0.70 to 0.90). The principal component 2 explained 17.6% of the total variation present in this Coimbra data set. For this PC2, the traits that

contribute the most for explaining variation were grain yield, ear number, ear length and number of kernel per row. The principal component 3 explained, in Coimbra, 12.7% of the total variation present in the data set. For this PC3, the trait contributing the most for explaining variation was the cob/ear weight per ear.

For Montemor, the principal component 1 explained 42.93% of the total variation present in the data set. For this PC1, the traits that contribute the most for the explaining variation were ear diameter 1 and 2 (correlation 0.90 to 1.00) and ear diameter 3 and 4, ear and kernel weight, cob diameters, medulla 1 and rachis (correlation, 0.70 to 0.90). The principal component 2 explained 17.09% of the total variation present in the Montemor data set. For this PC2, the trait contributing the most for the explaining variation was ear length. The principal component 3 explained, in Montemor, 11.06% of the total variation present in the data set.

VII.3.2 Map of the PB260 x PB266 progeny

The 17 AFLP primer combinations selected had a total of 451 dominantly scored polymorphic fragments on the F2 population, with an average of 27 polymorphic fragments per primer combination, ranging from 18 to 35, respectively, in the primer combinations E36-M49/E36-M50 and E32-M47. Among these 451, 227 were specific from PB260 and 224 were PB266 specific. In addition to the AFLP, 60 selected SSR markers were codominantly scored on the F2 population.

From the original molecular data set (149 F2 individuals screened with 511 markers—60 SSR and 451 AFLP polymorphic markers), we removed the 23 individuals and 57 markers with more than 10% missing values, plus 36 markers with a severe segregation distortion ($P \leq 0.05$) and 167 redundant markers clustered at the same position.

After performing a preliminary map analysis, three more individuals were removed due to their very improbable genotypes (singletons), as well as three more markers presenting suspected linkages with other markers.

Based on the remaining genotypic data of 248 markers screened on 123 F2 individuals, 10 linkage groups were obtained. Fifty-four markers were not assigned to any of the 10 resulting linkage groups. A linkage map containing 194 markers (144 dominant and 50 codominant) was developed, covering a total map distance of 1172.5 cM, with an average distance of approximately 6 cM per marker (Table VII.3).

Inspection of the individual linkage group $\chi 2$ values gave insights into the reliability of the obtained map. The $\chi 2$ values of the majority of the linkage groups were ≤ 1 except for linkage groups 1, 7 and 10, varying from 1.115 to 1.351 (Table VII.3). Given the high densities of markers, these $\chi 2$ values indicated that the map was relatively reliable. This map was then used for QTL identification.

Table VII.3 Refined genetic linkage map of the maize population F_2 (PB260xPB266) (used for QTL mapping)

Linkage Group	No. Markers	No. Codominant Markers	χ2 mean	Lenght (cM)	Average distance (cM)
1	29	6	1.115	167.63	5.78
2	27	7	1.060	146.00	5.41
3	19	5	0.924	145.16	7.64
4	22	3	0.792	107.96	4.91
5	18	6	0.598	152.78	8.49
6	14	4	0.724	85.72	6.12
7	16	5	1.351	93.45	5.84
8	16	5	1.098	90.17	5.64
9	16	4	0.913	75.67	4.73
10	17	5	1.141	107.93	6.35
Average	19.4	5	-	117.2	6.09
Total	194	50	-	1172.5	-

VII.3.3 QTL detected on the PB260 x PB266 progeny

Single trait QTL analysis. QTL were detected for the majority of the 29 traits with the exception of cob/ear weight at harvest, ear convulsion, kernel number, thousand kernel weight, kernels per row and ear weight at harvest. Sixty-five QTL (26 at the Coimbra environment and 39 at the Montemor environment), summarized in 17 different regions, were detected for 23 traits (Table VII.4). Eleven of these QTL were detected in both environments (constitutive) and distributed across four chromosomes (3, 5, 7 and 8). Strong clustering of QTL (with colocalized QTL for 3 or more traits) was observed in seven regions (Figure VII.1, Figure VII.2).

Four QTL detected for the fasciation trait were localized, one in chromosome 2 (Coimbra), and another in chromosome 10 (Montemor); two were constitutive in chromosome 7.

The amount of explained phenotypic variance ranged from 11.5% to 14.1% in each individual QTL detected, and in total the fasciation QTL explained 24.7% to 26.4% of the phenotypic variance at Coimbra and Montemor respectively (Figure VII.1, Table VII.4) assuming the absence of epistasis. In these detected fasciation QTL, the alleles for increasing the trait were always contributed by the parental accession PB260 (Table VII.4), with a lower level of fasciation, which is in agreement with the detected transgressive segregation in the F2 population (Table VII.2).

A single QTL was detected for yield in chromosome 6, accounting for 14.3% of the total phenotypic variance, and only at Montemor, where PB266 contributed with the allele increasing yield.

Two QTL were detected for ear number, in chromosomes 4 and 8, accounting for 16.7% and 12.1% of the total phenotypic variance at Montemor, respectively. In this case, the increasing alleles were, in the QTL located in chromosome 8 (en_m2), provided by PB266 and, in the QTL located in chromosome 4 (en_m1), by PB260 (Table VII.4).

Two QTL for ear length were constitutively detected in chromosomes 3 and 5. QTL on chromosome 3 explained 17.1% and 12.5% of the phenotypic variance at Coimbra and Montemor, respectively. The QTL in chromosome 5 explained 11.1% and 11.2% of the phenotypic variance (Figure VII.1, Table VII.4). Per environment, considering the absence of epistasis, the detected QTL explained a total of 28.0% (Coimbra) to 23.6% (Montemor) of the phenotypic variance for ear length.

Also in this trait and in agreement with the transgressive segregation detected in the F2 population, increasing alleles were contributed by the two parental lines (PB266 in the QTL detected in chromosome 5 and PB260 in the QTL detected in chromosome 3) (Table VII.2, Table VII.4).

Twenty-three QTL involved in inheritance of ear and cob diameters were detected, with a maximum of three QTL detected per trait in each environment. Ear and cob diameter QTL were mainly detected in chromosomes 1, 3 and 7, with chromosome 8 involved in cob diameter inheritance. QTL for cob diameter 1, in chromosome 8, and cob diameters 3 and 4, in chromosome 3, were constitutively detected.

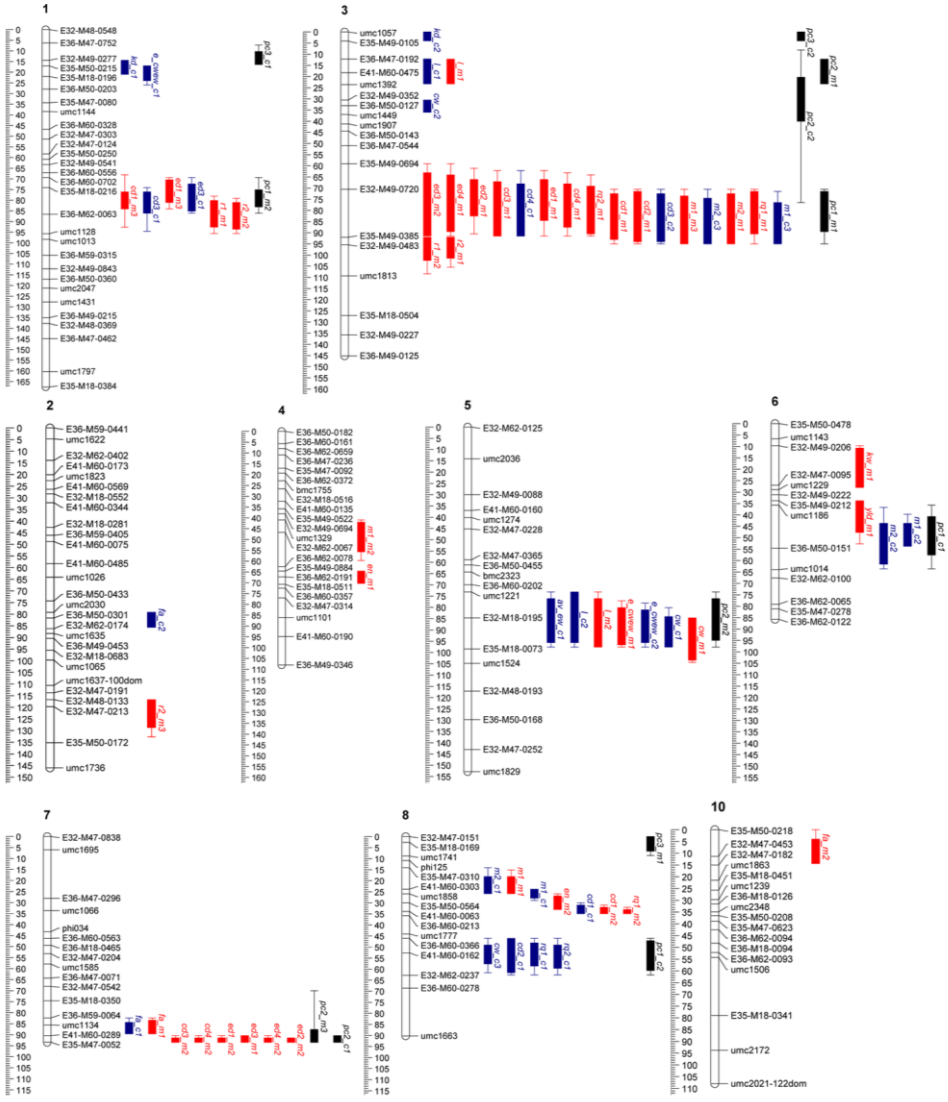

Figure VII.1 QTL detected for ear fasciation and related traits using 149 $F_{2:3}$ families (PB260 (non-fasciated) x PB266 (fasciated)) at two environments in Portugal (Coimbra and Montemor). Bar positions indicate the locations of quantitative trait *loci* (QTL): outer and inner interval correspond to 1-LOD and 2-LOD support interval, and are indicated as full box and a single line respectively. QTL nomenclature was arranged by the trait name plus environment abbreviation (c = Coimbra and m = Montemor) and the order number of detected QTL for the same trait in the genome (the higher the n°, the lower the contribution of the detected QTL for the explained phenotypic variability).

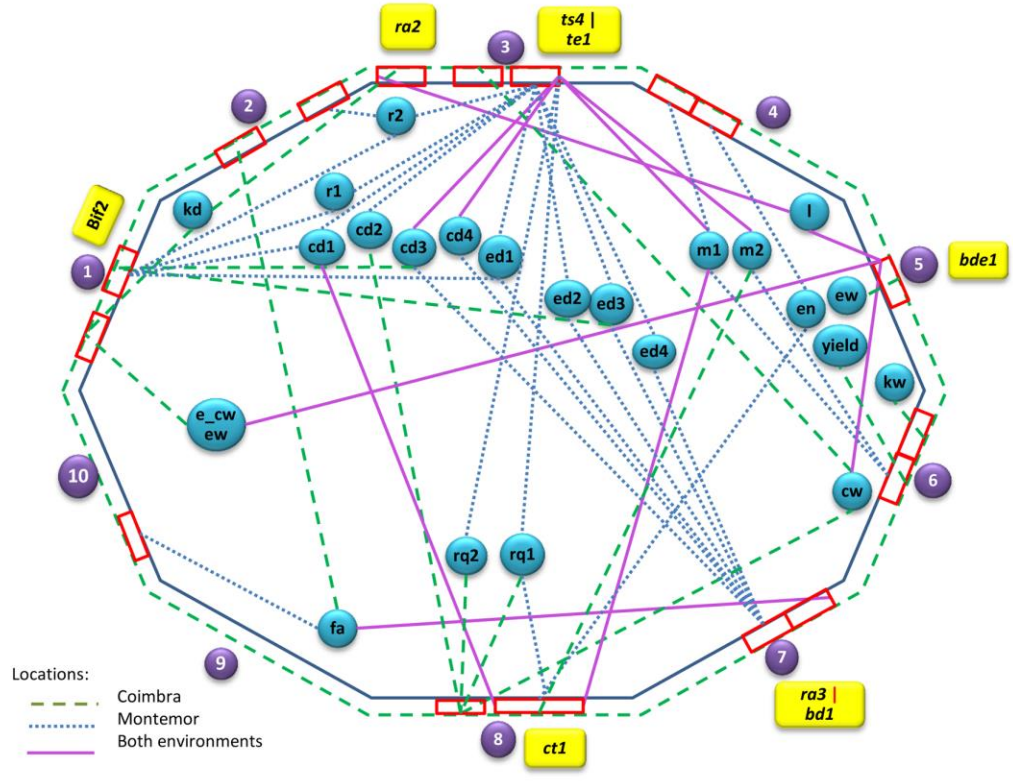

Figure VII.2 Representation of the ear fasciation and related traits QTL regions per maize chromosome, with indication of respective detection environment. For each chromosomal region, the respective candidate genes are indicated

Table VII.4 Quantitative Trait *Loci* for fasciation and related ear traits, estimated from 149 $F_{2:3}$ maize families (PB260 x PB266).

Trait	Environment[a]	QTL	Chromosome	QTL Position (cM)[b]	Flanking markers[c]	Peak LOD Score	Additive effect[d]	Gene Action[e]	R^{2f}
Yield	m	yld_m1	6	35.68	umc1186	4.12	-0.52	A	14.3
Ears number	m	en_m1	4	68.29	E36-M62-0191 / E35-M18-0511	5.34	0.90	PD	16.7
	m	en_m2	8	29.96	E35-M50-0564	3.97	-0.84	A	12.1
Ear lenght	c	l_c1	3	23.26	E41-M60-0475 / umc1392	5.69	0.86	PD	17.1
	c	l_c2	5	84.60	E32-M18-0195	3.83	-0.93	PD	11.2
	m	l_m1	3	15.24	E36-M47-0192 / E41-M60-0475	4.04	0.70	D	12.5
	m	l_m2	5	87.60	E32-M18-0195 / E35-M18-0073	3.61	-0.91	PD	11.1
Ear diameter 1	m	ed1_m1	3	74.47	E32-M49-0720 / E35-M49-0385	7.11	-0.25	PD	18.4
	m	ed1_m2	7	93.41	E41-M60-0289 / E35-M47-0052	4.50	0.09	OD	11.2
	m	ed1_m3	1	77.11	E35-M18-0216 / E36-M62-0063	3.62	-0.07	OD	8.7
Ear diameter 3	c	ed3_c1	1	79.11	E35-M18-0216 / E36-M62-0063	4.13	-0.20	PD	14.3
	m	ed3_m1	7	93.41	E41-M60-0289 / E35-M47-0052	5.64	0.11	OD	16.4
	m	ed3_m2	3	74.47	E32-M49-0720 / E35-M49-0385	5.04	-0.25	PD	14.4
Ear diameter 2	m	ed2_m1	3	73.47	E32-M49-0720 / E35-M49-0385	6.46	-0.23	D	19.1
	m	ed2_m2	7	93.45	E35-M47-0052	3.84	0.07	OD	10.9
Ear diameter 4	m	ed4_m1	3	74.47	E32-M49-0720 / E35-M49-0385	4.83	-0.21	PD	14.2
	m	ed4_m2	7	93.45	E35-M47-0052	4.62	0.06	OD	13.6
Kernel-row number 1	m	r1_m1	1	83.11	E35-M18-0216 / E36-M62-0063	5.89	-0.97	D	17.5
	m	r1_m2	3	95.01	E35-M49-0385 / E32-M49-0483	4.64	-1.05	PD	13.6
Kernel-row number 2	m	r2_m1	3	94.01	E35-M49-0385 / E32-M49-0483	6.64	-1.27	PD	17.2
	m	r2_m2	1	89.57	E36-M62-0063 / umc1128	6.50	-0.95	D	16.6
	m	r2_m3	2	121.65	E32-M47-0213 / E35-M50-0172	3.81	0.55	OD	9.3
Fasciation	c	fa_c1	7	86.57	umc1134 / E41-M60-0289	4.44	0.20	OD	13.2
	c	fa_c2	2	81.19	umc2030 / E36-M50-0301	3.89	0.31	PD	11.5

Trait	Environment[a]	QTL	Chromosome	QTL Position (cM)[b]	Flanking markers[c]	Peak LOD Score	Additive effect[d]	Gene Action[e]	R^{2f}
	m	fa_m1	7	87.57	umc1134 / E41-M60-0289	4.86	0.22	OD	14.1
	m	fa_m2	10	11.00	E35-M50-0218 / E32-M47-0453	4.22	0.30	PD	12.3
Kernel dept	c	kd_c1	1	19.02	E35-M50-0215 / E35-M18-0196	6.38	-0.04	A	18.5
	c	kd_c2	3	0.00	umc1057	4.81	-0.03	PD	13.6
Ear weight	c	ew_c1	5	84.60	E32-M18-0195	3.81	-15.09	PD	13.3
Kernel weight	m	kw_m1	6	21.69	E32-M49-0206 / E32-M47-0095	3.77	-9.43	A	13.2
Cob weight	c	cw_c1	5	91.60	E32-M18-0195 / E35-M18-0073	6.84	-4.86	A	16.8
	c	cw_c2	3	33.02	E36-M50-0127	4.31	3.28	PD	10.4
	c	cw_c3	8	52.64	E41-M60-0162	3.93	2.91	PD	9.4
	m	cw_m1	5	91.60	E32-M18-0195 / E35-M18-0073	4.88	-4.95	A	16.7
Cob/ear weight per ear	c	e_cwew_c1	1	21.02	E35-M50-0215 / E35-M18-0196	7.33	0.02	A	19.5
	c	e_cwew_c2	5	88.60	E32-M18-0195 / E35-M18-0073	5.99	-0.02	PD	15.3
	m	e_cwew_m1	5	87.60	E32-M18-0195 / E35-M18-0073	5.13	-0.02	PD	17.5
Cob diameter 1	c	cd1_c1	8	32.96	E35-M50-0564 / E41-M60-0063	5.38	0.16	PD	18.3
	m	cd1_m1	3	82.47	E32-M49-0720 / E35-M49-0385	5.84	-0.19	A	16.4
	m	cd1_m2	8	33.87	E41-M60-0063	5.45	0.15	PD	15.5
	m	cd1_m3	1	80.11	E35-M18-0216 / E36-M62-0063	3.78	-0.08	OD	9.0
Cob diameter 3	c	cd3_c1	1	81.11	E35-M18-0216 / E36-M62-0063	4.53	-0.13	D	13.8
	c	cd3_c2	3	83.47	E32-M49-0720 / E35-M49-0385	4.39	-0.14	PD	13.8
	m	cd3_m1	3	78.47	E32-M49-0720 / E35-M49-0385	6.30	-0.21	A	17.5
	m	cd3_m2	7	93.45	E35-M47-0052	6.09	0.10	OD	17.2
Cob diameter 2	c	cd2_c1	8	52.64	E41-M60-0162	3.77	0.12	PD	13.2
	m	cd2_m1	3	80.47	E32-M49-0720 / E35-M49-0385	5.07	-0.16	PD	17.3
Cob diameter 4	c	cd4_c1	3	79.47	E32-M49-0720 / E35-M49-0385	3.70	-0.13	PD	12.9
	m	cd4_m1	3	76.47	E32-M49-0720 / E35-M49-0385	6.29	-0.16	PD	18.2

Trait	Environment[a]	QTL	Chromosome	QTL Position (cM)[b]	Flanking markers[c]	Peak LOD Score	Additive effect[d]	Gene Action[e]	R^{2f}
	m	cd4_m2	7	93.45	E35-M47-0052	4.35	0.03	OD	12.3
Medulla 1	c	m1_c1	8	25.84	E41-M60-0303 / umc1858	7.10	0.09	PD	17.9
	c	m1_c2	6	51.68	umc1186 / E36-M50-0151	5.94	-0.10	D	13.3
	c	m1_c3	3	88.47	E32-M49-0720 / E35-M49-0385	4.01	-0.08	PD	9.5
	m	m1_m1	8	25.84	E41-M60-0303 / umc1858	6.26	0.11	PD	15.0
	m	m1_m2	4	47.66	E32-M62-0067 / E36-M62-0078	5.66	-0.12	A	13.7
	m	m1_m3	3	86.47	E32-M49-0720 / E35-M49-0385	4.71	-0.12	A	11.2
Medulla 2	c	m2_c1	8	23.84	E41-M60-0303	5.00	0.07	PD	12.8
	c	m2_c2	6	51.68	umc1186 / E36-M50-0151	4.43	-0.07	D	11.2
	c	m2_c3	3	86.47	E32-M49-0720 / E35-M49-0385	4.18	-0.07	PD	10.5
	m	m2_m1	3	82.47	E32-M49-0720 / E35-M49-0385	5.25	-0.12	A	17.8
Rachis 1	c	rq1_c1	8	52.64	E41-M60-0162	5.38	0.15	A	18.2
	m	rq1_m1	3	78.47	E32-M49-0720 / E35-M49-0385	5.77	-0.17	PD	16.5
	m	rq1_m2	8	34.87	E41-M60-0063 / E36-M60-0213	4.76	0.12	PD	13.7
Rachis 2	c	rq2_c1	8	52.64	E41-M60-0162	4.47	0.11	PD	15.4
	m	rq2_m1	3	79.47	E32-M49-0720 / E35-M49-0385	5.96	-0.15	PD	20.0

[a] m= Montemor; c= Coimbra
[b] QTL position in cM from the top of the chromosome
[c] molecular markers flanking the support interval estimated at a LOD fall of -2.00
[d] Additive effect = (phenotypic mean of the PB260 allele genotypes – phenotypic mean of the PB266 allele genotypes) / 2 [79]; negative values indicate that the PB266 allele increased trait additive value
[e] Gene action: A - additive if |dominant effect/additive effect| < 0.2; PD - partial dominance 0.2 < |d/a| < 0.8, D - dominance 0.8 < |d/a| < 1.2; OD - overdominance |d/a| > 1.2
[f] Percent explained phenotypic variance

The percentage of explained phenotypic variance per individual ear diameter QTL ranged from 8.7% in chromosome 1 (ear diameter 1 at Montemor) to 19.1% in chromosome 3 (ear diameter 2 at Montemor). For individual cob diameters QTL, the percentage of explained phenotypic variance ranged from 9.0% in chromosome 1 (Montemor) to 18.3% in chromosome 8 (Coimbra) (Figure VII.1, Table VII.4). Per environment, the total of explained phenotypic variance per ear diameters, considering the absence of epistasis, ranged from 14.3% (Coimbra, ear diameter 3) to 38.3% (Montemor, ear diameter 1) and per cob diameters from 12.9% (Coimbra, cob diameter 4) to 40.9% (Montemor, cob diameter 1). For cob and ear diameters QTL, 14 alleles increasing (chromosome 1 and 3, with 4 and 10 QTL respectively) and 9 alleles decreasing the traits (chromosome 7 and 8, with 6 and 3 QTL respectively; with chromosome 7 with QTL only detected at Montemor) were contributed by the parental accession PB266 (Table VII.2, Table VII.4).

QTL for row number 1 and 2 were only detected at one environment (Montemor) and included two colocalized QTL in chromosomes 1 and 3, for row number 1 and 2, and an additional QTL in chromosome 2, for row number 2. Individual QTL explained 13.6% to 17.5% of total phenotypic variance of kernel row number 1 (in total 31.1%, considering the absence of epistasis) and 9.3% to 17.2% in kernel row number 2 (in total 43.1%, considering the absence of epistasis) (Figure VII.1, Figure VII.2).

Four alleles for increasing the trait phenotypes (chromosome 1 and 3, with 2 QTL each) and one allele for decreasing (chromosome 2, exclusively for row number 2) were contributed by the parental accession PB266 (Table VII.2, Table VII.4).

Two QTL for kernel depth were detected in chromosome 1 and 3, only at one environment (Coimbra), explaining 13.6% to 18.5% of the total phenotypic variance, respectively. Increasing alleles were always contributed by PB266 (Figure VII.1, Table VII.4).

One QTL was detected for ear weight, only at Coimbra (explaining 13.3% of the phenotypic variance), in chromosome 5, where the increasing alleles were contributed by PB266. Also only one QTL was detected for kernel weight, at environment Montemor, in chromosome 6 (explaining 13.2% of the phenotypic variance) with the increasing allele being contributed by PB266 (Figure VII.1, Table VII.4).

One to three QTL were identified at Montemor and Coimbra, respectively, for cob weight in chromosome 3, 5 and 8. QTL detected in chromosome 5 were constitutive. Individual QTL explained 9.4% to 16.8% of the phenotypic variance and in total, per environment, explained 16.7% (Montemor) to 36.6% (Coimbra) of total phenotypic variance (Figure VII.1, Table VII.4). Both PB260 and PB266 contributed with alleles for increasing cob weight (Table VII.2).

For the cob/ear weight per ear, one to two QTL were detected per environment, one constitutive in chromosome 5 and another only detected in Coimbra (chromosome 1). The constitutive QTL

explained, respectively, 17.5% and 15.3% of the phenotypic variance in Coimbra and Montemor and the QTL at chromosome 1 for Coimbra, 19.5% of the phenotypic variance (Figure VII.1, Table VII.4). PB266 contributed with the increasing allele of the constitutive QTL detected in chromosome 5 and PB260 with the increasing allele of the QTL detected in chromosome 1 (Table VII.2, Table VII.4). In total, and per environment, the detected QTL explained 17.5% (Montemor) to 34.8% of total phenotypic variance of the cob/ear weight per ear.

For medulla 1 and 2, one to three QTL were detected, per environment, in chromosome 3, 4, 6 and 8. Two QTL for medulla 1 and 2 were constitutive, located, respectively, in chromosome 3 and 8. Individual QTL effects varied from 9.5% to 17.9% of explained variability. In total, and per environment, the detected QTL explained 39.9% (Montemor) to 40.7% (Coimbra) of the medulla 1 phenotypic variance, and 17.8% (Montemor) to 34.5% (Coimbra) of medulla 2, (Figure VII.1, Table VII.4).

For medulla QTL, 7 alleles for increasing the trait phenotypes (chromosome 3, 4 and 6, with 4, 1 and 2 QTL respectively) and 3 alleles for decreasing (chromosome 8) were contributed by the parental accession PB266, (Table VII.2, Table VII.4).

For rachis 1 and 2, one to two QTL were detected, per environment, in chromosome 3 and 8, but none of these QTL were constitutive. They explained a total of 18.2% (Coimbra) to 30.2% (Montemor) of the phenotypic variance of rachis 1 and a total of 15.4% (Coimbra) to

20.0% (Montemor) of rachis 2, considering the absence of epistasis (Figure VII.1, Table VII.4).

For rachis 1 and 2 QTL, alleles for increasing and for decreasing the traits were contributed by the parental accession PB266, (Figure III.1, Table VII.2, Table VII.6).

QTL analysis using principal components (PC). First three principal components accounting for 73% and 71% of variation in Caldeirão and Montemor, respectively, were used to map QTL associated with maize ear architecture (Figure VII.1, Table II.1).

At Coimbra, two QTL were detected for PC1. The first in chromosome 6 (13.4% of the phenotypic variance explained), colocalized with QTL for medulla1 and 2, strongly correlated traits and highly contributing to this PC1 vector. The second in chromosome 8 (13.1% of explained phenotypic variance) colocalized with QTL for rachis traits, cob diameter 2 and cob weight, traits that were moderate to very strongly correlated and highly contributing for PC1 in this environment. Additionally, in Coimbra, two QTL were identified for PC2, in chromosomes 3 and 7. The chromosome 3 PC2 QTL (12.8% of explained phenotypic variance) colocalized with QTL for length, cob weight, medulla, and cob diameters, with some of these traits strongly correlated (ear length with cob weight, cob diameter 3 with 4 and medulla 1 with 2). In chromosome 7, PC2 QTL (15.0% of explained phenotypic variance) colocalized with a QTL for ear fasciation. Finally, two QTL were detected in this environment for PC3 in chromosome 1 (18.95% phenotypic variance explained),

colocalizing with QTL for cob/ear weight per ear and kernel depth, traits that were weakly correlated, and in chromosome 3 (11.0% phenotypic variance explained), colocalizing with a kernel depth QTL detected on the same environment (Figure VII.1, Table VII.5).

At the Montemor environment, two QTL were identified for PC1. One QTL was located in chromosome 1 (12.2% phenotypic variance explained), colocalizing with QTL for row numbers, ear and cob diameter 1 (with row number 1 and 2 very strongly correlated), and another QTL in chromosome 3 (22.4% of the phenotypic variance explained), colocalizing with QTL for medulla and rachis traits, cob and ear diameters, with some of these traits very strongly correlated (ear diameter 1 and 2, ear diameter 3 and 4, cob diameter 1 and rachis 1, cob diameter 2 and rachis 2) (Figure VII.1, Table VII.5). Three PC2 QTL were detected for this environment in chromosomes 3, 5 and 7. In chromosome 3 (15.9% phenotypic variance explained), PC2 QTL colocalized with QTL for length, in chromosome 5 (12.4% of the phenotypic variance explained), with QTL for cob/ear weight per ear, cob weight and ear length, with some of these traits strongly correlated (cob weight with length and with cob/ear weight per ear), and in chromosome 7 (10.7% phenotypic variance explained), with QTL for all the ear diameters, cob diameter 3 and 4 and ear fasciation, with some of these traits strongly correlated (ear diameters 1 and 2, ear diameters 3 and 4, and fasciation with ear diameter 3). In the case of PC3, only one QTL was identified in chromosome 8 (explaining 15.3% of the phenotypic variance) and

localized away from any clustering QTL region (Figure VII.1, Table VII.5).

As expected, and as already highlighted for the PC QTL, many of the highly correlated individual traits presented colocalized QTL. Overall, based on QTL two LOD confidence intervals, all the QTL detected for the 23 individual measured traits were summarized as 17 different QTL clustered regions (Fig 1), seven of which had in common three or more traits. Within these seven highly clustered regions, three presented constitutive QTL. In particular, colocalization of constitutive QTL was observed in chromosome 3, among QTL for all the medulla traits and for cob diameters 3 and 4, with strong correlations detected in both environments between cob diameters 3 and 4 and between medulla 1 and 2. In chromosome 5, colocalization was detected among QTL for cob weight, ear length and cob/ear weight per ear. Strong correlations existed among these traits, except for length and cob/ear weight per ear for both locations. Finally, in chromosome 8, colocalization was detected between QTL for cob diameter 1 and medulla 1 and strong correlations existed between these two traits in both studied environments (Figure VII.1).

Table VII.5 Quantitative Trait *Loci* for the first three PCs, derived from 29 traits in the $F_{2:3}$ maize families (PB260xPB266) in two environments.

Environment	Principal component	QTL	Chromosome	QTL Position (cM)[a]	Flanking markers[b]	Peak LOD Score	Additive effect[c]	Gene Action[d]	R^2 [e]
Coimbra	PC1	pc1_c1	6	49.68	umc1186 / E36-M50-0151	4.35	-0.41	PD	13.4
		pc1_c2	8	52.64	E41-M60-0162	4.22	0.34	PD	13.1
	PC2	pc2_c1	7	93.41	E41-M60-0289 / E35-M47-0052	4.33	-0.13	OD	15.0
		pc2_c2	3	35.02	E36-M50-0127 / umc1449	4.34	0.21	PD	12.8
	PC3	pc3_c1	1	15.36	E32-M49-0277 / E35-M50-0215	6.37	-0.22	A	18.9
		pc3_c2	3	0.00	umc1057	3.89	-0.18	A	11.0
Montemor	PC1	pc1_m1	3	78.47	E32-M49-0720 / E35-M49-0385	7.63	-0.56	PD	22.4
		pc1_m2	1	79.11	E35-M18-0216 / E36-M62-0063	4.42	-0.23	OD	12.2
	PC2	pc2_m1	3	20.26	E41-M60-0475 / umc1392	5.43	0.22	PD	15.9
		pc2_m2	5	84.60	E32-M18-0195	4.31	-0.24	PD	12.4
		pc2_m3	7	91.41	E41-M60-0289 / E35-M47-0052	4.38	-0.14	OD	10.7
	PC3	pc3_m1	8	2.63	E35-M18-0169	4.43	-0.19	PD	15.3

[a] QTL position in cM from the top of the chromosome

[b] molecular markers flanking the support interval estimated at a LOD fall of -2.00

[c] Additive effect = (phenotypic mean of the PB260 allele genotypes – phenotypic mean of the PB266 allele genotypes) / 2 [79]; negative values indicate that the PB266 allele increased trait additive value

[d] Gene action: A - additive if |dominant effect/additive effect| < 0.2; PD - partial dominance 0.2 < |d/a| < 0.8, D - dominance 0.8 < |d/a| < 1.2; OD - overdominance |d/a| > 1.2

[e] Percent explained phenotypic variance.

VII.3.4 Putative candidate genes underlying detected QTL

From the 17 QTL regions, defined based on the QTL 2-LOD confidence intervals, we have selected eight different QTL regions to search for candidate genes (5 with constitutive QTL, plus 3 other with fasciation QTL or with QTL of fasciation highly correlated traits, ear and cob diameter 3 and row number 2) (Figure VII.2). On average, each region corresponded to 20.4 cM. Despite the exact physical distance

covered by these intervals being unknown, several candidate genes, mapping to the defined QTL regions confidence intervals, have been identified from the literature, based on their potential biological function (Table VII.6).

In the chromosome 1 region flanked by umc1144 and umc1128, where QTL for rows number 1 and 2, ear and cob diameter 1 (Montemor) and ear and cob diameter 3 (Coimbra) were colocalized, a possible candidate gene was *barren inflorescence 2* (*bif2*).The *bif2* is associated with maize architectural diversity and is known to affect the transition from inflorescence meristem to spikelet pair meristem or branch meristem (Upadyayula et al. 2006; Mcsteen P, Hake S 2001; Mcsteen et al. 2007; Pressoir et al. 2009). The *bif2* mutants have defects in the initiation of axillary meristems, and consequently, produce a reduced number of tassel branches and spikelets (Mcsteen P, Hake S 2001). Additionally, *bif2* mutants also produce a reduced number of ears, with fewer kernels, and their apical meristem is often fasciated (Skirpan et al. 2009). The *bif2* gene encodes a serine/threonine protein kinase that regulates polar transport of auxin (Mcsteen et al. 2007). BIF2 interacts with and phosphorylates BARREN STALK1 (BA1), a basic helix–loop–helix (bHLH) transcription factor required for axillary meristem initiation, suggesting that BA1 is a target of BIF2 (Gallavotti et al. 2004; Skirpan et al. 2008, 2009).

In this same region, several QTL for cob diameter (Veldboom, Lee 1994; Veldboom et al. 1994; Austin, Lee 1996) and kernel row

number 23 (Veldboom, Lee 1996a, b) were also previously detected (Table VII.6).

In the chromosome 3 region, ranging from umc1057 to umc1392, where the constitutive QTL for ear length was identified, we found *ra2* as a potential candidate gene. The *ra2*, similarly to *ra1*, has a highly branched and distorted ear, with irregular kernel placement (Vollbrecht et al. 2005). The *ra2* gene encodes a LOB (Bortiri et al. 2006) domain protein that determines the fate of stem cells in maize branch meristems (Bortiri et al. 2006; Gallavotti et al. 2010). The *ra2* regulates accumulation of *ra1* transcripts, placing the two genes in a single genetic pathway, with *ra2* upstream of *ra1* (Vollbrecht et al. 2005).

In the chromosome 3 region, ranging from umc1907 to umc1813, where the constitutive QTL for cob diameters 3 and 4 and medulla 1 and 2, plus ear and cob diameters, rachis and rows number for Montemor were identified, we found as a candidate gene the *tasselseed4* (*ts4*). The *ts4* encodes a mir172 microRNA that controls sex determination and meristem cell fate by targeting *Ts6/indeterminate spikelet1* (*ids1*) (Chuck et al. 2007). In addition, *ts4* not only affects sex determination, but can also cause inflorescence branching proliferation due to acquired indeterminacy of the spikelet pair meristem and the spikelet meristem (Mcsteen et al. 2000; Chuck et al. 2007; Vollbrecht, Schmidt 2009). The *ts4* mutants are characterized by irregular branching within the inflorescence and

feminization of the tassel caused by a lack of pistil abortion (Nickerson, Dale 1955).

Also in the same region of chromosome 3 (range umc1907 to umc1813), we found an additional potential candidate gene, *terminal ear1* (*te1*). Mutants of the maize *te1* gene have shortened internodes, abnormal phyllotaxy, leaf pattern defects and partial feminization of tassels. An earlike inflorescence forms in place of the normal terminal tassel. There is an increase in the frequency of leaf primordia initiation and the feminization of the terminal inflorescence on the main stalk (Veit et al. 1993; 1998). The *te1* gene encodes a RNA recognition motif (RRM) protein, and is expressed in the vegetative shoot apex, in semicircular rings, that laterally oppose the positions of leaf primordia (Veit et al. 1998).

In the chromosome 5 region, ranging from umc1221 to umc1524, where the constitutive QTL for cob weight, cob/ear weight per ear and ear length plus the QTL for ear weight, were detected only in Coimbra, the *bearded-ear1* (*bde1*) was indicated as a potential candidate gene. The *bde1* encodes zea agamous3 (zag3), a MADS box transcription factor belonging to the conserved AGAMOUS-LIKE6 clade (Thompson et al. 2009). The *bde1* is critical for multiple aspects of floral development, including floral meristem determinacy, organ development and sex determination. The *bde1* mutation affects floral development differently in the upper and lower meristem (Thompson et al. 2009). The upper floral meristem initiates extra

floral organs that are often mosaic or fused, while the lower floral meristem initiates additional floral meristems.

In the chromosome 7 region, ranging from umc1585 to the end of the chromosome, where QTL for fasciation, cob diameter 3 and related traits were identified, we have detected as a potential candidate gene *branched silkless1* (*bd1*) (Chuck et al. 2002), which encodes an ERF-like APETALA2 transcription factor and functions to repress indeterminate lateral branch meristem fates (Chuck et al. 2002).

The *bd1* was first described by Kempton (Kempton 1934). In mutants with strong alleles the ear spikelet meristems are replaced by branches similar to the tassel. In addition, no florets are initiated in the ear. In the tassel, the phenotype is less severe, possibly due to the expression of the duplicate of *bd1*. While the tassel spikelets are still indeterminate and branch-like, florets are initiated that produce viable pollen. The *bd1* is expressed in boundary domains, adjacent to the meristems that *bd1* also regulates. Phylogenetic analysis of the *bd1* gene demonstrated high conservation of the gene in all grass lineages with spikelets, indicating that the gene may be fundamental to spikelet initiation (Chuck et al. 2009; Monaco et al. 2013).

In the same QTL region we also found the *ra3* as a potential candidate gene. In its mutants, the axillary meristems can be enlarged and acquire abnormal identity or became indeterminate, leading to the production of long branches or more floral meristems in the ears. Tassels present the same developmental defects,

although at a lower frequency, leading to additional long branches (Monaco et al. 2013). The *ra3* encodes a functional trehalose-6-phosphate phosphatase, an enzyme that catalyzes the production of trehalose sugar and is expressed in discrete domains subtending axillary inflorescence meristems (Satoh-Nagasawa 2006). Genetic analysis has placed all three ramosa genes into a pathway, with *ra2* and *ra3* acting in parallel upstream of *ra1* (Satoh-Nagasawa 2006). It was proposed that RA3 regulates inflorescence branching by modification of a sugar signal that moves into axillary meristems. Alternatively, the fact that RA3 acts upstream of RA1 supports the hypothesis that RA3 itself may have a transcriptional regulatory function (Satoh-Nagasawa 2006).

Finally, in the chromosome 8 region, ranging from phi125 to umc1777, where the constitutive QTL for cob diameter 1 and medulla1 were located, we have identified as a potential candidate gene the ct1, whose mutant phenotype depicts semidwarf plants with furcated ears, but not fasciated, with all plant parts reduced proportionately (Jackson D, Hake S 2009).

Table VII.6 Candidate genes and previously described QTL for the currently detected ear fasciation and highly related traits QTL and other traits constitutive QTL. Main gene interactions and known homologies.

	BIN	1.04-1.07	3.02-3.04	3.05-3.09	5.04-5.06	7.02-end	8.03-8.05									
	Marker interval	**umc1144 - umc1128**	**umc1057-umc1392**	**umc1907-umc1813**	**umc1221-umc1524**	**umc1585-end**	**phi125-umc1777**									
	Detected QTL	ed3_c1 r2_m2 cd3_c1	l_c1 & m1	ed3_m2 cd3_c2 & m1 cd4_c1 & m1 m1_c3 & m3 m2_c3 & m1 r2_m1	l_c2 & m2 cw_c1 & m1 e_cwew_c2 & m1	ed3_m1 fa_c1 & m1 cd3_m2	cd1_c1 & m2 m1_c1 & m1	**Fasciation**	**Meristem**	**Ear**	**Tassel**	**Flowering**	**Tillering**	**Interaction**	**Homologue**	**Homology/Protein**
Candidate Gene/ QTL	*barren inflorescence2 (bif2)*	1.05								x	x		x	*ba1*		/serine threonine protein kinase
	QTL cob diameter 6	1.07								x						
	QTL kernel row 23	1.07								x						
	QTL cob diameter 12	1.07								x						
	QTL cob diameter 24	1.07								x						
	QTL cob diameter 28	1.07								x						
	ramosa2 (ra2)		3.02-3.03					x	x	x	x			*ra1*	*ra1, ra3*	LOB-domain TF/
	tasselseed4 (ts4)			3.04-3.05					x	x	x			*ids, ts6, sid1*		miR172/miR172 microRNA
	terminal ear1 (te1)			3.05				x		x	x				*ra1, ra3*	RNA-binding/RNA binding protein
	bearded-ear1 (bde1)				5.06				x	x		x				
	ramosa3 (ra3)					7.04		x	x	x				*fea1*	*ra1*	/Trehalose-6-phosphate phosphatase
	branched silkless1 (bd1)					7.04-7.06				x	x					AP2-domain TF
	compact plant1 (ct1)						8.01-8.03	x		x						

VII.4. Discussion

Fasciation is frequently found in the Portuguese maize germplasm (Vaz Patto et al. 2003). The knowledge of its genetic control could be used to better modulate yield while controlling the negative secondary effects of extreme fasciation expression (*e.g.* increasing yield, but maintaining uniformity of plants and ears) (Pêgo, Hallauer 1984). However, molecular genetic studies to understand the genetic basis of this trait were never performed on Portuguese germplasm.

To determine the genetic relationships among a comprehensive set of maize ear architecture traits related with fasciation, the current study presents a QTL analysis of the ear fasciation and related traits, for the first time undertaken on maize germplasm of Portuguese origin. This study also allowed us to propose potential candidate genes for ear fasciation. The results were obtained by repeated phenotypic analysis of the ear fasciation and related traits using a segregating F2 maize population of Portuguese origin (non-fasciated PB260 x fasciated PB266) that was also genotyped with AFLP and SSR markers.

QTL analysis revealed the existence of non-constitutive QTL for fasciation indicating a possible contribution of some minor environment-specific genes. However, also a limited experimental significance could be the cause for this non-detection considering our experimental limitations and the reduced number of environments tested. Even though polygenic (Pêgo, Hallauer 1984; Bommert et al. 2005, 2013), the inheritance of the ear fasciation trait in the

germplasm of Portuguese origin was not particularly complex (four QTL were detected for ear fasciation, one of them constitutive in the two studied environments), paving the way for relatively straightforward use of molecular markers in breeding programs by exploiting ear fasciation control.

In addition, 10 QTL were detected for the highest fasciation correlated traits (3 for ear diameter 3 and row number 2, and 4 in the case of cob diameter 3).

The number of detected QTL may have been underestimated, as QTL, may have escaped detection due to the scarcity of markers in some map regions, the small F2 population size (starting from 149 individuals) and a relatively high LOD threshold score used to reduce the rate of false positives (Young 1996). Future work should focus on the saturation of the genetic map presented here, with more codominant markers or other types of higher throughput dominant markers (such as Single Nucleotide Polymorphisms, SNP). This would allow gaps between distant markers to be filled, as well as increasing the likelihood of merging the total amount of screened markers into 10 linkage groups.

Even so, the presently detected QTL, considering the absence of epistasis and not assessing the variation potentially explained by QTL interactions could explain, per environment (Coimbra and Montemor), 24.7% and 26.4% of the phenotypic variation for ear fasciation; 14.3% and 30.8% of the phenotypic variation of ear diameter 3; 27.6% and 34.7% of the phenotypic variation of cob

diameter 3; and 43.1% of the phenotypic variation of row number 2 (although in this last case, these QTL were only detected at Montemor).

In the F2 population, a high correlation between ear fasciation and ear and cob diameter 3 and row number 2 was observed. This observation was consistent with the fact that the QTL for ear fasciation were colocalized, depending on the environment, at least with QTL for ear diameters 1 to 4 and cob diameters 3 to 4. QTL for ear fasciation and row number 2 were detected on the same chromosome (chromosome 2) but there was no overlapping of the respective confidence intervals. Nevertheless, taking into account the small population size, the non-overlapping of the two-LOD intervals of the particular ear fasciation and row number 2 QTL in chromosome 2 might have been caused by the choice of particular cofactor markers during the multiple QTL mapping approach.

The parental accession PB266 had an average higher level of ear fasciation than the parental accession PB260. Still, the alleles contributed by the parental accession PB266, in all the detected ear fasciation QTL including the constitutive QTL, decreased trait additive value. A similar situation occurred for the related trait row number 2, cob diameter 4, ear diameter 3 and 4 and cob/ear weight per ear detected QTL where, although the parental accession PB266 had significantly higher phenotypic values than the parental accession PB260, it contributed not only with alleles increasing, but also with alleles decreasing trait additive values. In fact, transgressive

segregation was observed for the majority of the analyzed traits, with phenotypic values of the two parental accessions significantly different from at least one of the two extremes of the F2 individuals' phenotypic range.

The knowledge of the genetic basis and location of the QTL responsible for the ear fasciation expression as well as its potential interaction with other related traits will facilitate the transfer of the milder fasciation alleles from the Portuguese germplasm to modern cultivars, hopefully without the negative effects of an extreme fasciation expression. As already highlighted, molecular markers could be used to support this introgression. In order to achieve this goal, the search for candidate genes as functional markers, arguably more promising and efficiently than the flanking markers for selection, is extremely important, providing that a direct association between function and phenotype exists.

The only constitutively detected fasciation QTL was located in chromosome 7, overlapping with QTL for ear and cob diameters, these last ones only detected in Montemor. Some of the possible candidate genes for fasciation in this position are *bd1* and *ra3* (mapped in the bin 7.04), which indeed can be related with the QTL for ear and cob diameters and fasciation identified in this chromosome 7 region (Table VII.6). The *bd1* affects ear branching architecture, being fundamental to spikelet initiation, and so could influence the ear fasciation or diameter traits presently studied. The *ra3* is of great value for ear architecture and establishes the correct

identity and determinacy of axillary meristems in both male and female inflorescences. However, previous studies with Portuguese maize fasciated germplasm, which did not consider the inbred lines that gave rise to the presently studied population, showed that the abnormal ear expression was not allelic to the three ramosa genes (Pêgo, Hallauer 1984). Yet in the present study *ra3* is seen as an interesting candidate gene. This support the existence of diversity in the genetic control of the fasciation expression among Portuguese maize germplasm, *i.e.*, different Portuguese germplasm may contain different combinations of different genes, all resulting in ear fasciation. In addition, the *ra2* was identified in the present study as a potential candidate gene, not for fasciation, but for ear length, in chromosome 3. Indeed, Pêgo and Hallauer studies (Pêgo, Hallauer 1984) stated that the genetic potential for increased yield in the fasciated Portuguese germplasm would be conditioned by the interaction between fasciation expression and ear length. In the present study, no strong negative correlation was detected between fasciation and length of the ear (-0.322 and -0,141 respectively for Coimbra and Montemor), probably due to the Portuguese farmers' selection criteria, which preferred long ears (although with fasciation expression). But these two traits are known to vary in opposite directions (Taguchi-Shiobara et al. 2001). In the present study we observed that the two parental accessions behave oppositely. PB260 was contributing positively to fasciation increase and negatively to ear length, while PB266 was contributing negatively to fasciation and

positively to ear length increase in the studied population (Table VII.2).

In the present study, other fasciation QTL were detected in chromosome 2 and 10, but not constitutively, indicating that these are also chromosomal regions that can be further explored to fully understand the genetic control of ear fasciation in the Portuguese germplasm. A QTL for row number 2 (only for Montemor) was also detected in this study in the chromosome 2 region, but in a different region from the fasciation QTL. In chromosome 10 QTL have been previously detected for grain yield (bin 10.03) (Ribaut et al. 1997; Schrag et al. 2006) and kernel row number per ear (bin 10.03– 10.07) (Ribaut et al. 2007). The ear fasciation trait had, in our study, a low correlation with yield; however, correlation with row numbers 1 and 2 (0.63 and 0.75, respectively) were much higher. Nevertheless, perhaps due to the lower resolution power of the present study, no constitutive kernel row number QTL was detected, and overall only for the Montemor environment did we detect kernel row number QTL in chromosomes 1, 2 and 3.

Also in chromosome 3, but away from the *ra2* location, QTL for ear diameter 3, row number 2 and cob diameter 3 (this one constitutively localized), were detected together with constitutive QTL for cob diameter 4 and medulla 1 and 2. Some of these traits had the highest correlation coefficients with fasciation (ear and cob diameters 3 and row number 2). In this region *ts4* and *te1* could be considered as potential candidate genes for the aforementioned fasciation- related

traits, due to their associated increased ear branching phenotypes, similar to a certain extent to what is found in the typical Portuguese maize fasciated traditional varieties. The *ts4* mutant phenotype presents a tassel compact silky mass, upright, with pistillate and staminate florets, with a proliferated, silky ear (Mcsteen et al. 2000; Vollbrecht, Schmidt 2009). In the *te1* mutant, kernel rows may be uneven and branches may form in the ear, depending on the allele and background (Weber et al. 2007).

Another constitutively detected QTL for ear length found in this study was located in chromosome 5, overlapping with the constitutively detected QTL for cob weight and cob/ear weight per ear. These traits were highly correlated with coefficients ranging from 0.69 to 0.98. A candidate gene in this chromosomal region might be the *bde1* (bin 5.06). Its mutants present polytypic and silky ears, showing a proliferation of pistillate tissue causing irregular growth on the ear and tassel. These phenotypes may indicate a possible influence of this gene on the traits for which QTL were detected in this position. Furthermore, in this same chromosome 5, the fea1 (Jackson et al. 2009) could also be indicated as an important candidate gene for these detected QTL, due to the small rounded ears and fasciated inflorescence meristems associated mutant phenotype, but its precise location is not yet known (Jackson et al. 2009).

In a chromosome 8 region (8.03–8.05), we constitutively detected two QTL for cob diameter 1 and medulla 1, traits that are strongly correlated. A possible candidate gene in this interval is the ct1, whose

mutant phenotype depicts semidwarf plants with furcated ears, but not fasciated, with all plant parts reduced proportionately (Jackson D, Hake S 2009). Furcated ears are very often observed among the Portuguese fasciated germplasm, with a strong effect on cob and medulla ear traits.

Other candidate genes related to ear fasciation, such as the *bif2*, were previously located in chromosome 1 (Mcsteen P, Hake S 2001), within the interval where we detected QTL for traits such as ear and cob diameter and row numbers, some of them highly correlated with ear fasciation (ear and cob diameter 3 and row number 2). The *bif2* mutants produce a reduced number of ears with fewer kernels. In addition, the apical meristem is often fasciated (Mcsteen P, Hake S 2001; Skirpan et al. 2008), a trait that is highly correlated with ear and cob diameters and row numbers. In this same region several ear traits QTL have also been previously detected by others. This is the case of the QTL cob diameter 6 (in bin 1.07) (Veldboom, Lee 1994; Veldboom et al. 1994) and QTL kernel row number 23 (Veldboom, Lee 1996a). Indeed, this chromosomal region appears to be highly associated with the inheritance of cob diameters, since other cob diameter QTL were also mentioned as near some of those internal markers (in bin 1.07), such as QTL for cob diameter 12 (Austin, Lee 1996), 24 and 28 (Veldboom, Lee 1996a).

In an attempt to identify the genomic regions controlling the most important factors contributing to the definition of the overall variation in maize ear architecture and yield, we detected QTL for the

Principal Components calculated separately for each environment. Colocalization of PC QTL and individual traits QTL was in accordance with the main contribution of each individual trait for each PC. Accordingly, in Coimbra, PC1 QTL overlapped individual QTL for cob weight, cob diameter, medulla and rachis in chromosomes 6 and 8; PC2 QTL overlapped individual QTL for ear length in chromosome 3 and, interestingly, also overlapped the constitutive ear fasciation QTL in chromosome 7, although fasciation is not one of the most contributing traits for this component. PC3 QTL overlapped individual QTL for kernel depth and cob/ear weight per ear in chromosome 1 and 3. In Montemor, PC1 QTL colocalized with individual QTL for cob and ear diameters in chromosome 1 and QTL for rachis, medulla and ear and cob diameters in chromosome 3. Also in chromosome 3, PC2 QTL overlapped with ear length QTL, and the same happened in chromosome 5. As already pointed out, Coimbra PC2 QTL also overlapped with the constitutive fasciation QTL in chromosome 7. Finally, the QTL detected for PC3 in Montemor did not overlap with any of the individual trait QTL in chromosome 8. Indeed for this principal component no individual trait contributed in a outstanding way. Possibly this QTL might be involved in a more overall regulation of multiple ear traits, which could not be detected using trait-by-trait analysis (Upadyayula et al. 2006). This fact reinforces the existence of recently detected regions that can be further explored in order to find new associations between QTL traits and candidate genes and to better understand and control fasciation in maize breeding.

Pêgo and Hallauer (Pêgo, Hallauer 1984) concluded that ear fasciation is a complex trait important in the Portuguese maize germplasm, with high potential for long-term maize breeding. In our study, since we used a segregating population developed from crossing only two contrasting inbreeds, we might have missed many of the alleles that control this ear trait in the Portuguese germplasm. In order to clarify which other genes or alleles are contributing to the fasciation expression in this interesting maize germplasm, future mapping approaches should consider multiparental populations or association mapping with an higher number of Portuguese-derived inbreed lines such as the ones described in Vaz Patto et al. (Vaz Patto et al. 2004).

In relation to the currently proposed candidate genes, fine mapping with additional markers in advanced Recombinant Inbred Lines (RIL) populations or complementary testing using Near Isogenic Lines (NIL) will be needed for the validation of some of the present hypothesis.

The present work represents the first molecular study in the elucidation of a set of genes controlling fasciation and associated molecular markers from the long-term legacy of Portuguese farmers, which after validation might have important breeding applications.

Portuguese farmers' selection of maize occurred over centuries and counted on the fasciation ear trait to increase ear size and yield. However, high levels of ear fasciation are associated with abnormal ear shapes that seriously limit harvesting. Additionally, its quantitative expression due to its genetic complexity and

dependency on environmental conditions hinders its current application in breeding programs. This is a very interesting trait for maize breeding, but one that must be fully understood at a genetic level before perfectly controlled in breeding programs. This control can be attained by the development of molecular selection tools based on QTL flanking molecular markers or associated functional markers (candidate genes), such as the ones identified in the present study. This study represents the first steps into the development of biotechnological tools for Marker Assisted Selection (MAS) of ear traits related to the typical fasciation of Portuguese maize germplasm. However prior to this, the newly identified QTL regions should be saturated with more molecular markers to increase the precision of QTL location and the linked flanking markers should be validated in other breeding populations. In our particular case a collection of diverse Portuguese maize inbred lines or the ear fasciation contrasting traditional maize landraces could be used to test if these trait /marker associations would be maintained in other genetic backgrounds. This breeding approach would ensure the use of a proper combination of genetic factors controlling ear diameter, kernel row number and ear length to allow ear fasciation expression without abnormal ear shapes and increasing yield and/or ear size, depending on the final breeding objective (Mendes-Moreira P 2008).

VII.5. Conclusions

We have detected significant variation for maize ear fasciation and related ear traits and mapped a number of QTL controlling those traits in the Portuguese derived PB260 x PB266 segregating population. We have found a substantial positive genetic correlation between ear fasciation and ear diameter 3, row number 2 and cob diameter 3, with heritabilities higher than 0.73. The constitutively detected QTL for fasciation was located in chromosome 7, indicating *ra3* as a putative candidate gene. This QTL mapping study has contributed to expanding the list of genomic areas involved in maize ear fasciation and related traits, especially in chromosomes 1, 3, 5, 7 and 8 where candidate genes *bif2*, *ra2*, *ts4*, *te1*, *bde1*, *ra3*, *bd1* and *ct1* and associated molecular markers were proposed.

VII.6. Material and Methods

VII.6.1 Population development

Based on information from the records of NUMI (a national maize breeding station in Braga, Portugal), two contrasting inbred lines for ear fasciation, PB260 (non-fasciated) and PB266 (fasciated) were selected as parental lines for the development of a fasciation segregating population. NUMI targets were the Portuguese farmers, who mainly used maize for bread production, *i.e.*, they selected mainly for white flint kernels with white cobs.

The PB260 pedigree is (PB6 x PB7) x PB6(2), PB6 being an inbred line derived from the Portuguese landrace 'Cem dias' and PB7 an inbred line derived from 'Northern White', an American population. PB260 was selfed for 19 years. From three years of field evaluations, PB260 presented an average of 72 days for male and female flowering, and in a scale from 1 to 4 (where 1 is the minimum and 4 the maximum), 1.3 for vigor, 2 for plant height, 4 for uniformity of the plants in the plot, 2 for plant lodging and 1.2 for Sesamia spp. resistance. The ear height insertion was 3, in a scale from 1 to 9 (where 1 is the minimum and 9 the maximum and 5 corresponds to the middle of the plant). The ear shape was conical, with white flint kernel type and white cob, with a fasciation level of 1.41 in a scale from 1 to 9 (where 1 is the minimum and 9 the maximum).

The pedigree of PB266, also known at NUMI as WF9R, is (WF9 x PB53) x WF9. The WF9 is a yellow dent inbred line originally selected by the Indiana Agriculture Experimental Station from the population Wilson Farm Reid (USDA & ARS-GRIN 2013). Historically, the name of WF9 was kept at NUMI, although when introduced into the Portuguese breeding program in the 40s, this yellow dent kernel line, as many others, was converted to a white inbreed line by crossing with Portuguese germplasm. In particular, this conversion included crosses with Portuguese germplasm with white abnormal ears, followed by several backcrosses to the recurrent parents (Pêgo, Hallauer 1984). The PB53 was derived from 'Northern White', an American population (Runge et al. 2004).

The PB266 line was also selfed for 19 years, before being used in this study. PB266 is characterized by 74 days for male and female flowering, and in the same 1 to 4 scale described before, it presents 1.2 for vigor, 3 for plant height, 4 for uniformity of the plants in the plot, 3 for plant lodging and 3.5 for Sesamia spp resistance. Following the 1 to 9 scale, its ear height insertion was also 3. The ear shape was conical, with white dent kernel type and white cob, with a fasciation level of 2.38. Due to the relatively small ear fasciation differences among PB260 and PB266, this cross allowed us to identify genes contributing to a milder ear fasciation expression variation.

Vaz Patto et al. (Vaz Patto et al. 2004) studying the genetic diversity of a collection of Portuguese maize inbred lines, clustered PB260 together with white flints of a Portuguese origin. PB266 was not analyzed in that study however, its genetic distance, later computed, to WF9 was 0.197, while to PB260 it was 1.063 (Alves ML, unpublished results), which indicates PB266 clustering nearby WF9, on the yellow dent germplasm group of American origin, and away from PB260. Indeed PB266 was selected to be the Portuguese WF9 version, *i.e.*, with white kernel and cob, and with an early cycle more adapted to the national farming systems and more resistant to *Sesamia* spp.

PB260 and PB266 were crossed to develop an F1 hybrid. A F1 hybrid plant was self-pollinated to obtain an F2 population. 149 randomly chosen F2 plants were selfed to obtain 149 $F_{2:3}$ families. Leaf samples were collected from each of the 149 F2 plants for DNA extraction and

molecular markers analysis. The $F_{2:3}$ families derived from the 149 F2 individuals were used to evaluate ear fasciation and related ear architecture traits. The evaluation occurred under field and laboratory conditions.

VII.6.2 Field experiments and phenotypic evaluations

The 149 $F_{2:3}$ families were evaluated at two environments in Portugal (Coimbra 40°13'0.22"N, 8°26'47.69"W and Montemor 40°10'4.82"N, 8°41'14.84"W) in 2008. These two environments are a part of the Mondego irrigation perimeter, a very high-yielding area where the average yield for maize hybrids is 14.5 $Mgha^{-1}$. Montemor is located 21 km from the sea coast and Coimbra 50 km. Both environments have an altitude of 25m. Both Montemor and Coimbra have alluvial soils, but compared to Montemor, Coimbra has a lower soil pH (5.2 *versus* 6.3) and a lower percentage of soil with a particle size less than 0.2 mm diameter (86.9% *versus* 92.5%); it also has a higher percentage of organic matter (2.3% *versus* 1.7%). The agricultural practices were similar in both environments; however the sowing date in 2008 was May 9 at Coimbra and May 28 at Montemor and the harvest from October 2 and 21, respectively.

In each environment, a randomized complete block design, with two replications, was used. Each plot consisted of one single row with 3.1 m (2.6 m planted row plus 0.5 m, space between two planted rows) long, with an inter-row distance of 0.75 m. Each plot was overplanted by hand and thinned at the V7 growth developmental stage (Ritchie

et al. 1993) for a stand of approximately 50000 plants ha^{-1}. Plots were mechanically and/or hand-weeded and managed following common agricultural practices for maize in the region. All the plots were harvested by hand.

Phenotypic data were collected for 29 traits and are described in Table VII.1. Some traits were measured per plot (traits 1–4, Table VII.1), such as grain yield ($Mgha^{-1}$) adjusted to 15% grain moisture at harvest. All the other traits were measured on five ears per plot, randomly selected after harvest, and dried (35°C) to approximately 15% grain moisture to ensure *ceteribus paribus* conditions during measurements. Following this procedure, 25 measurements were made per ear (traits 5–29, Table VII.1) (Mendes-Moreira et al. 2014). The five ears average value per plot was considered for data analysis (Table VII.1).

VII.6.3 Statistical analysis of phenotypic data

Phenotypic data descriptive statistics were calculated using SAS (the SAS system for Windows, version 9.2, Cary, NC, USA). Pearson's correlation coefficients were computed for each trait between environments as well as between all traits by PROC CORR procedure. The distributions of the traits in each environment were tested using the Kolmogorov-Smirnov's test of normality. PROC GLM procedure was used for analysis of variance. Environments (Coimbra and Montemor) and genotypes were treated as fixed effects. Repetitions, treated as random, were nested in the environments. Genotype x Environment interaction was included in the model. The PROC

VARCOMP was used to estimate variance components for each trait in each environment separately as well as for both environments. Broad-sense heritabilities, representing the part of the phenotypic variance in the total phenotypic variance, were calculated for each environment as: h2 = Vg 2/ [Vg 2 + (V2/r)], where Vg 2 is the genotypic variance, V2 is the error variance and r is the number of replications, and for both environments as: h2 = Vg 2/ [Vg 2 + (Vge 2/e) + (V2/re)], where Vge 2 is the G x E interaction variance and e is the number of environments.

In order to have an indication of possible transgressive segregation among parental lines and the $F_{2:3}$ families, we compared the average data of PB260 and PB266 obtained at Coimbra during 2010 and 2012 field trials, with two repetitions in organic production, with the average extremes of the $F_{2:3}$ families field trials (obtained as described in field experiments and phenotypic evaluations section). For the extremes of the $F_{2:3}$ families we considered the top five maximum and top five minimum values per trait. Analysis of variance was applied to these data. When significant differences were detected, the Shéffe test was used to compare parental and extreme F2 averages (Table VII.2).

Principal component analysis (PCA) was performed using PROC PRINCOMP procedure in SAS considering all phenotypic traits separately for each environment in order to isolate the most important factors contributing to the definition of the overall variation in maize ear architecture and yield. The first three principal

components were used to map QTL associated with the overall variation of maize ear architecture and yield in a multivariate approach.

VII.6.4 Marker analysis and linkage map construction

Plant leaf samples were collected from the 149 F2 PB260xPB266 individuals at V9 to V12 stages of growth and development, from the self-pollination field trial. These 149 F2 self-pollinated individuals produced sufficient seed for establishing the $F_{2:3}$ families' multilocation trials.

DNA was extracted from each F2 plant leaf sample using a modified CTAB procedure (Saghai-Maroof et al. 1984). The F2 population individuals were analyzed using Simple Sequence Repeats (SSRs) and Amplified Fragment Length Polymorphism (AFLP) markers.

SSR protocol. The SSR marker technique was performed as described by Vaz Patto et al. (2004) with minor modifications. Forward primers of SSR primer pairs were labeled with two fluorescence dyes (IRDye 700 or IRDye 800) (Eurofins MWG Operon, Germany) to allow amplification fragments analysis using a 4300 DNA analyzer system (LI-COR Biosciences, USA). SSR alleles were detected and scored using SAGA Generation 2 software (LI-COR Biosciences, USA).

In order to select the most informative SSR primer pairs, the parental lines, PB260 and PB266, and a F1 individual were screened with 211 SSR markers chosen from Maize Genetics and Genomic Database

(MaizeGDB) (Lawrence et al. 2008) based on their repeat unit and bin location. This resulted in the selection of 60 SSR primer pairs that were amplified on the F2 individuals. Primer sequences are available from the MaizeGDB. The amplification fragments size was determined in base pairs and visually scored (peak detection) at least twice independently for each entry, to ensure data accuracy. Data were recorded as present (1), absent (0) or missing (-), allowing the construction of a binary matrix of the SSRs phenotypes.

AFLP Protocol. The AFLP technique was performed using the AFLP Analysis System I (Invitrogen, Carlsbad, CA, USA) kit protocol, with minor modifications. The EcoRI primers were labeled with two fluorescence dyes (IRDye 700 or IRDye 800) (Eurofins MWG Operon, Germany) to allow amplification fragments analysis using an 4300 DNA analyzer system (LI-COR Biosciences, USA). MseI primers with only two selective nucleotides were also tested to increase the total number of amplified fragments per primer combination. The primer core sequences were those of Vos et al. (1995). Thirty-six EcoRI/MseI base primer combinations were first tested in the parental lines (PB260 and PB266) in order to select the most informative primer combinations. This resulted in the selection of 17 different primer combinations that were used to screen the 149 F2 individuals.

Clearly readable amplified fragments of the 149 accessions were determined for size in base pairs and visually scored at least twice independently for each entry; they were recorded as present (1), absent (0) or missing (-)(USDA & ARS-GRIN 2013).

This allowed the construction of the binary matrix of the different AFLP phenotypes. This matrix, together with the SSR data matrix, was used for the construction of the input file for JoinMap 4.0 software (Van Ooijen 2006).

Map construction. Linkage analysis and segregation distortion tests ($P \leq 0.05$) were performed using JoinMap 4.0 software (Van Ooijen 2006). The determination of linkage groups of markers was done with a LOD score of 3. The linkage map calculations were done using all pairwise recombination estimates lower than 0.49 and a LOD score higher than 0.01, and applying the Kosambi mapping function (Kosambi 1944).

Individuals and markers with more than 10% missing values were removed from the original molecular data set. Also markers with a severe segregation distortion ($P \leq 0.005$) were excluded.

After a preliminary map analysis, improbable genotypes, including double recombination events (singletons), markers with suspected linkages with other markers and redundant markers clustered at the same position were removed, following the approach of Vaz Patto et al. (Vaz Patto et al. 2007). All of the codominant markers were kept in this refined map. To check the reliability of the obtained map, the individual linkage group $\chi 2$ was inspected.

Linkage groups were assigned to the corresponding chromosome using the SSR map locations from the consensus maize map as in the Maize Genetics and Genomics Database, MaizeGDB (Monaco et al. 2013). This was also a check for the accuracy of the composition of

linkage groups, as only markers assigned to the same chromosome should be present in the same linkage group.

VII.6.5 QTL analysis

The previously obtained F2 refined linkage map was used for QTL identification. Kruskal-Wallis single-marker analysis (non-parametric test), as well as for both interval mapping (Lander, Botstein 1989) and multiple-QTL mapping (MQM) (Jansen, Stam 1994) were performed using MapQTL version 4.0 (Van Ooijen 2002). A backward elimination procedure was applied to select cofactors significantly associated with each trait at $P < 0.02$ to be used in MQM. Genome-wide threshold values ($P < 0.05$) for declaring the presence of QTL were estimated from 10,000 permutations of each phenotypic trait (Churchill, Doerge 1994). The 1-LOD and 2-LOD support intervals were determined for each LOD peak.

The R^2 value, representing the percentage of the phenotypic variance explained by the marker genotype at the QTL, was taken from the peak QTL position as estimated by MapQTL. Additive and dominance effects for detected QTL were estimated using the MQM procedure. Gene action was determined following Stuber et al. (Stuber et al. 1987) as: additive (d/a = 0–0.20); partial dominance (d/a = 0.21–0.80); dominance (d/a = 0.81–1.20); and overdominance (d/a > 1.20), where, d/a = dominance effects/additive effects. Maps were drawn using MapChart version 2.2 software (Voorrips 2002). QTL analysis was performed on entry means from individual environments. The QTL nomenclature corresponded to the trait's abbreviation (Table

VII.1) followed by the environment abbreviation (c = Coimbra and m = Montemor), and finally a rank number, indicating the contribution of the QTL for a certain trait (based on R2) (Table VII.4).

QTL for different traits were declared as potential ‘‘common QTL” when they showed overlapping confidence intervals (Tian et al. 2011). On the other hand, “constitutive QTL” referred to a stable QTL across both environments (Collins et al. 2008).

Potential candidate genes and previously published QTL were identified for the ear fasciation and highly related traits and for all the constitutive QTL regions. This search was performed by comparing the 2-LOD confidence interval positions of the presently detected QTL with the known locations of genes and QTL affecting yield and ear architecture traits at the consensus Maize IBM2 2008 Neighbours Frame Map, available from MaizeGDB (Monaco et al. 2013). The presently detected 2-LOD confidence interval SSR flanking markers were used as anchor markers in these map comparisons. The qTeller toolbox (Schnable, Freeling 2011) was also helpful on the QTL position comparisons.

VII.7. Acknowledgements

Conception: MCVP

Design of the work: MCVP

Acquisition of data: phenotypic (PMM, MA, JPPS, JPNS, JCS) and molecular (MA, PMM)

Analysis and interpretation of data: PMM, MA, ZS, MCVP

Article drafting: PMM, MCVP

Revising it critically: ARH, SP, MCVP, ZS, MA

Populations' development and breeding: SP, PMM

VII.8. References

Austin DF, Lee M (1996) Comparative mapping in $F_{2:3}$ and F6:7 generations of quantitative trait *loci* for grain yield and yield components in maize. Theor Appl Genet 92: 817-826. doi: 10.1007/BF00221893 PMID: 24166546

Bai F, Reinheimer R, Durantini D, Kellogg EA, Schmidt RJ (2012) TCP transcription factor, BRANCH ANGLE DEFECTIVE 1 (BAD1), is required for normal tassel branch angle formation in maize. Proc Natl Acad Sci USA 109:12225-12230. doi: 10.1073/pnas.1202439109 PMID: 22773815

Beavis WD, Grant D, Albertsen M, Fincher R (1991) Quantitative trait *loci* for plant height in four maize populations and their associations with qualitative genetic *loci*. Theor Appl Genet 83: 141-145. doi: 10. 1007/BF00226242 PMID: 24202349

Bommert P, Lunde C, Nardmann J, Vollbrecht E, Running M, Jackson, D, Hake, S, Werr W (2005) thick tassel dwarf1 encodes a putative maize ortholog of the Arabidopsis CLAVATA1 leucine-rich repeat receptor-like kinase. Development 132:1235-1245. PMID: 15716347

Bommert P, Nagasawa NS, Jackson D (2013) Quantitative variation in maize kernel row number is controlled by the FASCIATED EAR2 locus. Nat Genet 45:334–337. doi: 10.1038/ng.2534 PMID: 23377180

Bortiri E, Chuck G, Vollbrecht E, Rocheford T, Martienssen R, Hake S (2006) *ramosa2* encodes a LATERAL ORGAN BOUNDARY domain protein that determines the fate of stem cells in branch meristems of maize. Plant Cell 18: 574-585. PMID: 16399802

Brewbaker JL (2009) Double-cob (dbcb) on chromosome 1. Maize Genetics Cooperation Newsletter 1

Brown PJ, Upadyayula N, Mahone GS, Tian F, Bradbury PJ, Myles S, Holland JB, Flint-Garcia S, McMullen MD, Buckler ES (2011) Distinct genetic

architectures for male and female inflorescence traits of maize. PLoS Genetics doi: 10.1371/journal.pgen.1002383.

Carillo P, Feil R, Gibon Y, Satoh-Nagasawa N, Jackson, D, Bläsing OE, Stitt M, Lunn JE (2013) A fluorometric assay for trehalose in the picomole range. Plant Methods 9: 21. doi: 10.1186/1746-4811-9-21 PMID: 23786766

Chuck G, Candela H, Hake S (2009) Big impacts by small RNAs in plant development. Curr Opin Plant Biol 12: 81-86. doi: 10.1016/j.pbi.2008.09.008 PMID: 18980858

Chuck G, Meeley R, Hake S (2008) Floral meristem initiation and meristem cell fate are regulated by the maize AP2 genes *ids1* and *sid1*. Development 135:3013-3019. doi: 10.1242/dev.024273 PMID: 18701544

Chuck G, Meeley R, Irish E, Sakai H, Hake S (2007) The maize tasselseed4 microRNA controls sex determination and meristem cell fate by targeting Tasselseed6/indeterminate spikelet1. Nat Genet 39: 1517-1521. PMID: 18026103

Chuck G, Muszynski M, Kellogg E, Hake S, Schmidt RJ (2002) The control of spikelet meristem identity by the branched silkless1 gene in maize. Science 298(5596): 1238-1241. PMID: 12424380

Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963-971. PMID: 7851788

Collins NC, Tardieu F, Tuberosa R (2008) Quantitative trait *loci* and crop performance under abiotic stress: where do we stand? Plant Physiol 147: 469-486. doi: 10.1104/pp.108.118117 PMID: 18524878

Eckardt NA (2007) Evolution of compound leaf development in legumes: Evidence for overlapping roles of KNOX1 and FLO/LFY genes. The Plant Cell Online 19: 3315-3316. doi: http://dx.doi.org/10.1105/tpc.107.057497

Emerson RA (1912) Inheritance of certain “abnormalities” in maize. Am Breed Assoc Rept 8: 385–399

Eveland AL, Satoh-Nagasawa N, Goldshmidt A, Meyer S, Beatty M, Sakai H, Ware D, Jackson D (2010) Digital gene expression signatures for maize development. Plant Physiol 154: 1024-1039. doi: 10.1104/pp.110. 159673 PMID: 20833728

Ferrão JEM (1992) A aventura das plantas e os descobrimentos portugueses. Programa Nacional de Edições Comemorativas dos Descobrimentos Portugueses, Portugal

Gallavotti A, Long JA, Stanfield S, Yang X, Jackson D, Vollbrecht E, Schmidt RJ (2010) The control of axillary meristem fate in the maize ramosa pathway. Development 137: 2849-2856. doi: 10.1242/dev. 051748 PMID: 20699296

Gallavotti A, Zhao Q, Kyozuka J, Meeley RB, Ritter MK, Doebley JF, Pe ME, Schmidt RJ (2004) The role of barren stalk1 in the architecture of maize. Nature 432: 630-635. PMID: 15577912

Hayes HK (1939). Linkage relations of gl4 with wx and sh. Maize Genet Coop News Lett. 13: 1–2.

Holland JB, Coles ND (2011) QTL controlling masculinization of ear tips in a maize (*Zea mays* L.) intraspecific cross. G3 (Bethesda) 1:337-341. doi: 10.1534/g3.111.000786 PMID: 22384344

IBPGR (1991) Descriptors for maize, Mexico City. International Board for Plant Genetic Resources, Rome

IPGRI (2000) Descritores para o milho, Mexico City. International Plant Genetic Resources Institute, Rome

Irish EE, Langdale JA, Nelson TM (1994) Interactions between tassel seed genes and other sex determining genes in maize. Dev Genet 15: 155-171

Jackson D (2009) Vegetative Shoot Meristems. In: Bennetzen JL, Hake SC, editors. Handbook of Maize: Its Biology: Springer Science.

Jackson D, Hake S (2009) The genetics of ear fasciation in maize. MNL 2

Jackson DP, Nagasawa NS, Nagasawa N, Sakai H (2009) Nucleotide sequences encoding RAMOSA3 and sister of RAMOSA3 and methods of use for same. Google Patents. http://www.google.com/patents/US20060191040. Accessed 18 April 2014

Jansen RC, Stam P (1994) High resolution of quantitative traits into multiple *loci* via interval mapping. Genetics 136: 1447-1455. PMID: 8013917

Jiao Y, Zhao H, Ren L, Song W, Zeng B, Guo J, Wang B, Liu Z, Chen J, Li W (2012) Genome-wide genetic changes during modern breeding of maize. Nat Genet 44:812-815. doi: 10.1038/ng.2312 PMID: 22660547

Kempton J (1934) Heritable characters in maize XLVII—branched silkless. J Hered 25: 29-32

Kempton JH (1923) Heritable characters of maize XIV—branched ears. J Hered 14: 243-251

Kosambi DD (1944) The estimation of map distances from recombination. Ann Eugenics 12: 172–175

Lander ES, Botstein D (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185-199. PMID: 2563713

Lawrence, CJ, Harper, LC, Schaeffer, ML, Sen, TZ, Seigfried, TE, Campbell, DA (2008) MaizeGDB: The maize model organism database for basic, translational, and applied research. Int J Plant Genomics 2008: 496957

Liu R, Jia H, Cao X, Huang J, Li F, Tao Y, Qiu F, Zheng Y, Zhang Z (2012) Fine mapping and candidate gene prediction of a pleiotropic quantitative trait locus for yield-related trait in *Zea mays*. PloS ONE doi: 10.1371/journal.pone.0049836

Lu Y, Yan J, Guimaraes CT, Taba S, Hao Z, Gao S, Chen S, Li J, Zhang S, Vivek BS, Magorokosho C, Mugo S, Makumbi D, Parentoni SN, Shah T, Rong T, Crouch JH, Xu Y (2009) Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. Theor Appl Genet 120: 93-115. doi: 10.1007/s00122-009-1162-7 PMID: 19823800

Matthews D, Grogan C, Manchester C (1974) Terminal ear mutant of maize (*Zea mays* L.) J Agric Sci 82: 433-435. doi: http://dx.doi.org/10.1017/S0021859600051327

Mcsteen P, Hake S (2001) barren inflorescence2 regulates axillary meristem development in the maize inflorescence. Development 128: 2881-2891. PMID: 11532912

Mcsteen P, Laudencia-Chingcuanco D, Colasanti J (2000) A floret by any other name: control of meristem identity in maize. Trends Plant Sci 5: 61-66. PMID: 10664615

Mcsteen P, Malcomber S, Skirpan A, Wu X, Kellogg E, Hake S (2007) barren inflorescence2 encodes a co-ortholog of the PINOID serine/threonine kinase and is required for organogenesis during inflorescence and vegetative development in maize. Plant Physiol 144:1000-1011. PMID: 10664615

Mena M, Mandel MA, Lerner DR, Yanofsky MF, Schmidt RJ (1995) A characterization of the MADS-box gene family in maize. Plant J 8:845-854. PMID: 8580958

Mendes-Moreira P, Pêgo S, Vaz Patto MC, Hallauer A (2008) Comparison of selection methods on 'Pigarro', a Portuguese improved maize population with fasciation expression. Euphytica 163: 481-499. doi: 10.1007/s10681-008-9683-8

Mendes-Moreira PM, Mendes-Moreira J, Fernandes A, Andrade E, Hallauer AR, Pêgo SE, Vaz Patto M (2014) Is ear value an effective indicator for maize yield evaluation? Field Crop Res 161: 75-86. doi: 10.1016/j.fcr.2014.02.015

Modarres A, Dijak M, Mather D, Smith D, Hamilton R, Dwyer L, Stewart D (1998) Leafy reduced-stature maize hybrid response to plant population density and planting patterns in a short growing season area [*Zea mays* L.-Quebec (Canada)]. Maydica 43: 227-234. doi: 10.2135/cropsci1999.0011183X003900040025x

Monaco MK, Sen TZ, Dharmawardhana PD, Ren L, Schaeffer M, Naithani S, Amarasinghe V, Thomason J, Harper L, Gardiner J (2013) Maize metabolic network construction and transcriptome analysis. The Plant Genome doi: 10.3835/plantgenome2009.01.0001. doi: 10.3835/ plantgenome2009.01.0001

Nelson O, Ohlrogge A (1957) Differential responses to population pressures by normal and dwarf lines of maize. Science 125: 1200-1200. PMID: 13432785

Nelson OE, Ohlrogge A (1961) Effect of heterosis on the response of compact strains of maize to population pressures. Agronomy J 53: 208-209

Neuffer MG, Coe EH, Wessler SR (1997) Mutants of maize, Cold Spring Harbor Laboratory Press, NY.

Nickerson NH, Dale EE (1955) Tassel modifications in *Zea mays*. Ann Missouri Bot Gard 42: 195-211

Pautler M, Tanaka W, Hirano HY, Jackson D (2013) Grass Meristems I: Shoot apical meristem maintenance, axillary meristem determinacy, and the floral transition. Plant Cell Physiol 54: 302-12. doi: 10. 1093/pcp/pct025 PMID: 23411664

Pêgo S (1982) Genetic potential of Portuguese maize with abnormal ear shape, Ph.D. Thesis, Iowa State Univ.

Pêgo SE, Hallauer AR (1984) Portuguese maize germplasm with abnormal ear shape. Maydica 29: 39–53

Phipps I (1928) Heritable characters in maize. J Hered 19: 399-404

Pressoir G, Brown PJ, Zhu W, Upadyayula N, Rocheford T, Buckler ES, Kresovich S (2009) Natural variation in maize architecture is mediated by allelic differences at the PINOID co-ortholog barren inflorescence2. Plant J 58: 618–628. doi: 10.1111/j.1365-313X.2009.03802.x PMID: 19154226

Ribaut JM, Fracheboud Y, Monneveux P, Banziger M, Vargas M, Jiang C (2007) Quantitative trait *loci* for yield and correlated traits under high and low soil nitrogen conditions in tropical maize. Mol Breed 20: 15-29. doi: 10.1007/s11032-006-9041-2

Ribaut JM, Jiang C, Gonzalez-De-Leon D, Edmeades G, Hoisington D (1997) Identification of quantitative trait *loci* under drought conditions in tropical maize. 2. Yield components and marker-assisted selection strategies. Theor Appl Genet 94: 887‑896

Ritchie SW, Hanway JJ, Benson GO (1993) How a corn plant develops. Iowa State University CES Special Report 48: 21 pp

Runge EC, Smith CW, Betrán J. Corn: origin, history, technology, and production: Wiley. Com; 2004.

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW. 1984. Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci 81: 8014-8018. doi: /10.1073/pnas.81.24.8014

Satoh-Nagasawa N, Nagasawa N, Malcomber S, Sakai H, Jackson D (2006) A trehalose metabolic enzyme controls inflorescence architecture in maize. Nature 441: 227-230. PMID: 16688177. doi:10.1038/nature04725

Schnable JC, Freeling M (2011) Genes identified by visible mutant phenotypes show increased bias toward one of two subgenomes of maize. PLoS ONE 6(3): e17855. doi: 10.1371/journal.pone.0017855 PMID: 21423772

Schrag TA, Melchinger AE, Sorensen AP, Frisch M (2006) Prediction of single-cross hybrid performance for grain yield and grain dry matter content in maize using AFLP markers associated with QTL. Theor Appl Genet 113: 1037-47. PMID: 16896712

Skirpan A, Culler AH, Gallavotti A, Jackson D, Cohen JD, McSteen P (2009) BARREN INFLORESCENCE2 interaction with ZmPIN1a suggests a role in auxin transport during maize inflorescence development. Plant Cell Physiol 50: 652-657. doi: 10.1093/pcp/pcp006 PMID: 19153156

Skirpan A, Wu X, McSteen P (2008) Genetic and physical interaction suggest that BARREN STALK1 is a target of BARREN INFLORESCENCE2 in maize inflorescence development. Plant J 55: 787-797. doi: 10.1111/j.1365-313X.2008.03546.x PMID: 18466309

Steinhoff J, Liu W, Reif J, Porta G, Ranc N, Würschum T (2012) Detection of QTL for flowering time in multiple families of elite maize. Theor Appl Genet 125: 1539-1551. doi: 10.1007/s00122-012-1933-4 PMID: 22801873

Stuber CW, Edwards M, Wendel J (1987) Molecular marker-facilitated investigations of quantitative trait *loci* in maize. II. Factors influencing yield and its component traits. Crop Sci 27: 639-648

Taguchi-Shiobara F, Yuan Z, Hake S, Jackson D (2001) The *fasciated ear2* gene encodes a leucine-rich repeat receptor-like protein that regulates shoot meristem proliferation in maize. Genes Dev 15: 2755-2821. PMID: 11641280

Thompson BE, Bartling L, Whipple C, Hall DH, Sakai H, Schmidt R, Hake S (2009) bearded-ear encodes a MADS box transcription factor critical for maize floral development. Plant Cell 21: 2578-2590. doi: 10.1105/tpc.109.067751 PMID: 19749152

Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, Mcmullen MD, Holland JB, Buckler ES (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet 43:159-162. doi:10.1038/ng.746

Upadyayula N, Da Silva H, Bohn M, Rocheford T (2006) Genetic and QTL analysis of maize tassel and ear inflorescence architecture. Theor Appl Genet 112: 592-606. PMID: 16395569. doi: 10.1007/s00122-005-0133-x

USDA & ARS-GRIN (2013) Wf9, Ames 19293 - *Zea mays* subsp. *mays* - Wf9. USDA, National Plant Germplasm System, Germplasm Resources Information Network - (GRIN). http://www.ars-grin.gov/cgi-bin/npgs/acc/display.pl?1082719http://www.ars-grin.gov/cgi-bin/npgs/acc/display.pl?1082719. Accessed 19 April 2014

Van Ooijen J (2006) JoinMap 4. Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, Netherlands

Van Ooijen J, Boer M, Jansen R, Maliepaard C (2002) Map QTL 4. Plant Research International. Wageningen, the Netherlands

Vaz Patto MC, Moreira PM, Carvalho V, Pêgo S (2007) Collecting maize (*Zea mays* L. convar. *mays*) with potential technological ability for bread making in Portugal. Genet Res Crop Evol 54:1555‑1563. doi: 10.1007/s10722-006-9168-3

Vaz Patto MC, Rubiales D, Martin A, Hernandez P, Lindhout P, Niks R, Stam P (2003) QTL mapping provides evidence for lack of association of the

avoidance of leaf rust in Hordeum chilense with stomata density. Theor Appl Genet 106: 1283‑1292. PMID: 12748780

Vaz Patto MC, Šatović Z, Pêgo S, Fevereiro P (2004) Assessing the genetic diversity of Portuguese maize germplasm using microsatellite markers. Euphytica 137: 63-72. doi:10.1023/B:EUPH.0000040503.48448.97

Veit B, Briggs SP, Schmidt RJ, Yanofsky MF, Hake S (1998) Regulation of leaf initiation by the terminal ear 1 gene of maize. Nature 393: 166-168. PMID: 9603518

Veit B, Schmidt RJ, Hake S, Yanofsky MF (1993) Maize floral development: New genes and old mutants. Plant Cell 5: 1205-1215. PMID: 12271023

Veldboom LR, Lee M (1994) Molecular-marker-facilitated studies of morphological traits in maize. II: Determination of QTLs for grain yield and yield components. Theor Appl Genet 89: 451-458. doi: 10. 1007/BF00225380 PMID: 24177894

Veldboom LR, Lee M (1996a) Genetic mapping of Quantitative Trait *Loci* in maize in stress and nonstress environments: I. Grain yield and yield components. Crop Sci 36: 1310-1319. doi:10.2135/cropsci1996.0011183X003600050040x

Veldboom LR, Lee M (1996b) Genetic mapping of Quantitative Trait *Loci* in maize in stress and nonstress environments: II. Plant height and flowering. Crop Sci 36: 1320-1327. doi:10.2135/cropsci1996.0011183X003600050041x

Veldboom LR, Lee M, Woodman WL (1994) Molecular marker-facilitated studies in an elite maize population: I. Linkage analysis and determination of QTL for morphological traits. Theor Appl Genet 88: 7-16. doi: 10.1007/BF00222387 PMID: 24185875

Vollbrecht E, Schmidt RJ (2009) Development of the Inflorescences. In: Bennetzen JL, Hake SC (eds) Handbook of maize: its biology. Springer, pp 13‑40

Vollbrecht E, Springer PS, Goh L, Buckler ES, Martienssen R (2005) Architecture of floral branch systems in maize and related grasses. Nature 436: 1119-1126. PMID: 16041362. doi:10.1038/nature03892

Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. J Hered 93: 77-78. PMID: 12011185. doi: 10.1093/jhered/93.1.77

Vos P, Hogers R, Bleeker M, Reijans M, Van De Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, et al. (1995) AFLP: a new technique for DNA

fingerprinting. Nucl Acids Res 23:4407-4407. PMID: 7501463. doi: 10.1093/nar/23.21.4407

Weber A, Clark RM, Vaughn L, De Jesus Sánchez-Gonzalez J, Yu J, Yandell BS, Bradbury P, Doebley J (2007) Major regulatory genes in maize contribute to standing variation in teosinte (*Zea mays* ssp. *parviglumis*). Genetics 177: 2349-2359. PMID: 17947410. doi: 10.1534/genetics.107.080424

Weber AL, Briggs WH, Rucker J, Baltazar BM, De Jesus Sanchez-Gonzalez J, Feng P, Buckler ES, Doebley J (2008) The genetic architecture of complex traits in teosinte (*Zea mays* ssp. *parviglumis*): New evidence from association mapping. Genetics 180: 1221-1232. doi: 10.1534/genetics.108.090134 PMID: 18791250

White OE (1948) Fasciation. Bot Rev 14: 319-358

Wu X, Skirpan A, McSteen P (2009) Suppressor of sessile spikelets1 functions in the ramosa pathway controlling meristem determinacy in maize. Plant Physiol 149: 205-219. doi: 10.1104/pp.108.125005 PMID: 18997117

Yamasaki M, Tenaillon MI, Vroh Bi I, Schroeder SG, Sanchez-Villeda H, Doebley JF, Gaut BS, McMullen MD (2005) A large-scale screen for artificial selection in maize identifies candidate agronomic *loci* for domestication and crop improvement. Plant Cell 17: 2859-2872. PMID: 16227451. doi: http://dx.doi.org/10.1105/tpc.105.037242

Young N (1996) QTL mapping and quantitative disease resistance in plants. Annu Rev Phytopathol 34: 479-501. PMID: 15012553. doi: 10.1146/annurev.phyto.34.1.479

Zhang G, Wang X, Wang B, Tian Y, Li M, Nie Y, Peng Q, Wang Z (2013) Fine mapping a major QTL for kernel number per row under different phosphorus regimes in maize (*Zea mays* L.). Theor Appl Genet 126: 1545-1553. doi: 10.1007/s00122-013-2072-2 PMID: 23494393

Zhang H, Zheng Z, Liu X, Li Z, He C, Liu D, Luo Y, Zhang G, Tan Z, Li R (2010) QTL mapping for ear length and ear diameter under different nitrogen regimes in maize. Afr J Agric Res 5: 626‑630. doi: 10.1111/j.1439-0523.2007.01465.x

Zhou ML, Zhang Q, Sun ZM, Chen LH, Liu BX, Zhang KX, Zhu XM, Shao JR, Tang YX, Wu YM (2013) Trehalose metabolism-related genes in maize. J Plant Growth Regul. doi:10.1007/s00344-013-9368-y

CHAPTER VIII.

General Discussion

INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
/UNL

Knowledge Creation

VIII.1. Overall discussion

VIII.1. Overall discussion

Plant breeding is one of the corner stones to solve the next world challenges, the feeding 9,000 million people with a growing competition for land, water and energy. This scenario, with climatic changes as a background, is complemented with the need to reduce the impact of the food system on the environment and on the human health. This complexity needs a transdisciplinary approach that also includes sustainability, biodiversity, genetics, molecular, statistics, economics, participatory research and information technologies.

Towards knowledge integration, the overall aims of this study was to evaluate the participatory maize breeding evolution under the VASO project *via* phenotypic and molecular data. This study was complemented with the development of a formula that could be useful for farmers' selection in a PPB methodology towards yield increase, and with the genetic basis elucidation of the ear trait fasciation, a very important ear trait to PPB farmers as a way to maintain the population resilience and yield enhancement.

Since its beginning in 1984, VASO project used an integrative philosophy defined by Pêgo (Pêgo, Antunes 1997). In 2006 an overview on the opportunities that VASO project create for genetic diversity conservation and breeding was published as part of this PhD thesis (Chapter II). These opportunities included the adaptation to marginal areas of production, to sustainable agriculture and integrating local knowledge. Posterior characterization of the long term participatory plant breeding work of VASO project included a

detailed phenotypical and molecular analysis (Chapters III, IV and V). Phenotypic data, beyond its direct information, could be of special interest as farmers' selection tools, providing knowledge to farmers on selection procedures. With this purpose we improved the existent "ear value formula" as a farmer's selection tool to increase yield based on ear traits (Chapter VI). In addition a more detailed genetic study was performed for ear fasciation, a quantitative trait that has been continuously selected by Portuguese farmers', and despite its morphological variation, with an effective impact on yield (Chapter VII).

VIII.1.1 Evaluation of the long term participatory plant breeding

VASO as a long term participatory plant breeding project (PPB), has contributed simultaneously for conservation of genetic resources and landraces improvement implicitly oriented for maize bread (the majority of the farmers involved on this project used maize to produce bread maize). In our study, a quantitative approach, using 'Pigarro' (a white flint Portuguese maize landrace) (Chapter IV and V) and 'Fandango' (a yellow dent, Portuguese maize synthetic population with high degree of fasciation) (Chapter III), allowed a comparative evaluation of the applied farmer's and breeder's selection methods. 20 years of farmer's mass selection and 12 years of breeder's S2 lines recurrent selection were used for 'Pigarro' improvement. In addition, 19 years of mass selection partially done

both by the breeder (until cycle 5) and by the farmer (after cycle 5) were applied to 'Fandango' improvement. Comparisons were based on several years of field trials, with a detailed agronomic and ear morphological characterization. Additionally, for 'Pigarro', the genetic diversity evolution was evaluated along cycles and between selection methods, using molecular markers (Chapter V).

Farmer and breeder selection had different goals. The farmer aimed at the maximization of ear size, while the breeder aimed at increasing yield and uniformity of some traits (*e.g.* ear height).

Farmer's selection response analysis using mass selection both on 'Pigarro' and 'Fandango' (after cycle 5) indicated that the plant life cycle increased and ears became shorter and larger, with more and smaller kernels. In addition, 'Pigarro' tassels became bigger and ears increased their row numbers, becaming more fasciated and with more convulsions. In the case of 'Fandango' plant and ear height became higher.

On the breeders' selection for 'Pigarro' (by S2 lines), the plants became smaller, ears became thinner, with less kernel row numbers, fasciation and moisture. A yield decrease was also observed. In case of 'Fandango' breeder's selection (using mass selection until the cycle 5) a slight increase in kernels per row was observed.

According the results obtained, it was observed for 'Fandango' that the yield gain per cycle and per year was 3.09% for Lousada (the environment where the selection took place) *versus* 0.63% for all other trial locations. These differences indicate that long term

selection of ‘Fandango’ was effective for local adaptation. This result suggest a study at molecular level to analise the variation of the number of alleles during selection, as done for ‘Pigarro’ (Chapter V). In addition a more detailed study on the genetic composition changes of ‘Fandango’, along selection, is also suggested regarding its 76 inbred lines background.

The comparison between breeder selection and farmer selection indicated that ears phenotypically changed particularly under farmer selection. These changes were depicted by an increase on ear and cob diameters, rachis, kernel row number, convulsion and fasciation with a tendency for a decrease in ear length. In summary farmers selected for shorter and wider ears, with increased levels of fasciation and smaller kernels. In the case of breeder selection, ears became longer and less fasciated, with an overall increase of crop uniformity.

These results showed that mass selection (with a 1–5% of selection pressure) was more effective for increasing yield than S2 lines recurrent selection (with a 15–20% of selection pressure). However with S2 lines recurrent selection, and in case of ‘Pigarro’ a more uniform population was obtained which fulfils some of the breeding programs requests. In the case of ‘Fandango’, uniformity was maintained, but plant became taller. For both selection methods, cobs have become wider and heavier with progress of the selection cycles. Both selection methods maintained the ability for

polycropping systems and quality for bread production according to Vaz Patto et al. (2009, 2013).

The yield decrease for both 'Pigarro' (S2 lines recurrent selection) and 'Fandango' (after C5) can be explained by the low effectiveness of selection due to the exclusion of stalk lodged plants in the basic units of selection. Considering that Hallauer and Sears (1969) observed that in the absence of a correlation between yield and stalk lodging, the exclusion of stalk lodged plants reduces the intensity of selection for yield from 7.5 to 27.4%.

It was also observed that across the selection cycles of 'Fandango', the area needed per plant became higher, *i.e.*, plants needed lower plant densities to produce ears. In our trials (with a fixed plant density) competition for space was more severe in advanced cycles and some plants did not yield any ear. Additionally, for 'Pigarro' S2 lines recurrent selection, yield decrease could be related with fasciation expression decrease and also to the mentioned exclusion of stalk and root lodged plants during selection. Because we are combining selection for yield with root and stalk lodging it may require additional cycles of recurrent selection. Hallauer et al. (2010) describes three cycles of recurrent selection for first brood European corn borer and stalk rot resistance developed populations having acceptable levels of resistance. Indicating that if we combine selection for these traits with selection for yield, some trade-offs usually are made in the final selections for recombination. Instead of three cycles, it may require three to six cycles of selection to attain a

comparable level of resistance. Progress will be made but at lower rate because of the compromises made in the selection process.

In addition to the phenotypic characterization, also a molecular characterization was used for 'Pigarro' both on farmers and breeder's selection cycles. During both selection approaches, genetic diversity changed, to allow population to phenotypically respond to selection, but was not reduced even with the most intensive breeder's selection, maintaining the necessary resilience to further adapt (Vaz Patto et al. 2008). Our molecular diversity evolution analysis emphasized potential associations between particular neutral molecular markers and the *loci* controlling some of the phenotypic traits under selection (*e.g.*, ear length, fasciation and related ear traits as ear diameter and kernel-row number) (Chapter IV and V). These associations need however to be better explored and validated by future linkage or association mapping approaches previous to their use for supporting trait selection in sustainable farming systems (Chapter VI and VII).

Both farmer's and breeder's selection methods were effective for diversity conservation, but their choice will depend on maize breeding program aims: Phenotypic recurrent selection is easier and potentially cheaper to adopt by farmers for OPV (Open Pollinated Varieties) improvement, whereas breeder selection results in a more uniform crop, being more adapted to hybrid development programs.

VIII.1.2 Farmers' selection tools

Across farmer's and breeder's long term selection, several phenotypical changes are observed in the traits expression. The monitoring of these changes are essential to better understand, both farmers' and breeders' selection procedures and to improve selection indexes such as the ear value formula. Ear value (EV) formula was developed in 1993 under the scope of a Portuguese regional maize ear competition (the "Sousa Valley Best Ear Competition"). EV formula included ear length, kernel weight at 15% moisture, number of rows and number of kernels per ear. This formula had two main purposes, ears evaluation for the ear competition and maize selection for breeding. EV formula was based on published maize trait correlations, with no direct inputs from farmers maize yield. To fulfill this gap we improved this formula analyzing in detail a set of populations where the best Sousa Valley ears came from. This data set analyzed represented a broad range of plants and ears. Data analyses helped us to identify, what were the major components that explain a complex trait such as yield. Yield is an expression of fitness and radical changes in one yield component are accompanied by adjustments in other component(s), implying the existence of correlated gene frequencies changes. This fact explains that the same yield increase can be obtained selecting for different traits combinations and originating different phenotypes (*e.g.* bigger ear *versus* prolificacy, prolificacy *versus* higher densities). In the case of 'Fandango' and 'Pigarro' using correlations and analysis methods as MARS (multiple adaptive regression splines) (Friedman 1991), CART

(Classification and Regression Trees) (Breiman et al. 1984) and RF (Random Forest) (Breiman 2001) we identify ear weight, kernel depth and rachis 2 as the most important traits related with yield, followed by cob and ear diameters and kernels per row. These data were obtained from the representativeness of the traits regarding the six best ranked methods, excluding fixed traits models. The maize plant density in the field (stand) was the most important field variable related with yield. However, it was not used for this maize regional competition.

Later on we use the best formula to predict yield and to test the quality of the prediction using different interpretable methods. With exception of the first method (mars.ears), composed by 12 of the 23 measured traits, the following four ranked methods obtained had only 4 variables or terms. To compare these more simple models we developed and applied a ranking method, latter refined by Ribeiro de Brito (2014). The selected formula was entitled EVA formula (Ear Value Adjusted formula). EVA formula showed to be the best compromise solution due to a reduced complexity when compared to the other models, exclusion of field traits, and easy to use regarding the number and the traits used. The EVA formula traits indicated that kernel weight, ear length, kernel row number and number of kernels are the most important traits to be use both for farmers' and breeders's selection and for providing a better ears ranking for the "Sousa Valley best ear competition. Due to its simplicity, EVA formula can be easily adopted by farmers and associations interested in

germplasm conservation and development. Besides, a smaller number of traits is less expensive to measure.

The use of the EVA formula on the maize ear competition had contributed to connect local knowledge and scientific knowledge under a collaborative research approach. Consequently this formula can be used as a tool in Participatory Plant Breeding (PPB) projects where quantitative information is collected and used by farmers to improve their own selection procedures. The empirically derived models in this study were specific to the range of the populations used in the competition of —Sousa Valley Best Ear. To some extent, such models can be calibrated for use with other maize populations. Furthermore these models can be used for pre-breeding, on-farm conservation, organic and low input agriculture, polycropping systems *i.e.*, germplasm adaptation to different environments. This is possible because this models can indicate us the most indicated traits that can help us to select the best germplasm for a certain environment or production system, providing also the knowledge for adaptation to climate changes. These models can also elucidate the breeding selection procedures evolution along time and allow comparing the work among breeders. Another application of this models, can use data set from UPOV test guidelines across years or according the breeder. This allows to search, what were the traits related with yield and what were the phenotypic changes across time (*e.g.* ideotype, upright leaves) and according breeders.

Finally the EVA formula can be a starting point for a more active long term engagement of farmers with germplasm development and an open door to a better understanding of quantitative genetics by farmers (Chapter VI).

VIII.1.3 Fasciation

Knowledge of the genes affecting maize ear inflorescence traits may lead to better grain yield modeling. Maize ear fasciation, defined as abnormal flattened ears with high kernel row number, is a quantitative trait widely present in Portuguese maize landraces. Maize fasciation has also attract more recently the interest of scientists due to its potential relation to an increased yield (Allen et al. 2013; Jackson et al. 2011; Pautler et al. 2015). Portuguese farmers' must have been interested on this ear fasciation trait since maize introduction in the country, considering its presence on old traditional landraces. This farmers' interest influenced Portuguese breeders', and brought this trait into national breeding programs (*e.g.* the NUMI hybrid, "HB19") and to the participatory OPV breeding program VASO.

Phenotypic studies on this ear trait were precluded with Portuguese germplasm (Pêgo 1982; Pêgo, Hallauer 1984), however no molecular studies existed before the present study. To fulfill this gap an $F_{2:3}$ population, was developed from a cross between contrasting inbred lines (non fasciated PB260 x fasciated PB266) towards the elucidation of the genetics of the fasciation trait. We have detected significant

variation among parental inbred lines PB260 and PB266, and respective minimum and maximum of the $F_{2:3}$ families for maize ear fasciation and related ear traits. With this study we mapped a number of QTLs controlling those traits in the Portuguese derived PB260 x PB266 segregating population. We have found a substantial positive correlation between ear fasciation and ear diameter 3, row number 2 and cob diameter 3, with heritabilities higher than 0.73. The constitutive QTL detected for fasciation was located in chromosome 7, indicating *ramosa3* (*ra3*) as a putative candidate gene. In addition, this QTL mapping study has contributed to expand the list of genomic areas potentially involved in maize ear fasciation and related traits, especially in chromosomes 1, 3, 5, 7 and 8 where other candidate genes *barren inflorescence2* (*bif2*), *ramosa2* (*ra2*), *tasselseed4* (*ts4*), *terminal ear1* (*te1*), *bearded-ear1* (*bde1*), *branched silkless1* (*bd1*) and *compact plant1* (*ct1*) were proposed, with flanking selecting neutral molecular markers.

In case of 'Pigarro', potential associations between particular neutral molecular markers and the *loci* controlling some of the phenotypic traits under selection (*e.g.*, ear length, fasciation and related ear traits as ear diameter and kernel-row number) were detected by a molecular evolution analysis. We found that some of the associations detected for 'Pigarro' occurred also in the segregating PB260 x PB266 population for umc1907, umc1524 and umc1858. The umc1907, on the bin3.05, was significantly out of Hardy-Weinberg equilibrium ($P< 0.05$) in all selection cycles (farmer's and breeder's) associated

with fasciation decrease and cycle duration for breeder and increase for farmer's selection, with *te1* as candidate gene. The umc1524 on bin 5.06, was associated with a decrease of tassel and the ear height until the second cycle of breeder selection and an increase in kernel weight from the second to the third breeding cycle. The *bde1* as candidate gene related with multiple aspects of floral development including floral meristem determinacy, organ development and sex determination is probably related with the phenotypic traits observed (Chapter IV, V and VII).

VIII.1.4 Key findings and advances

The phenotypic and molecular evaluation of the VASO project long term participatory maize breeding work, highlight the following aspects:

1) Phenotypic recurrent selection (farmer's selection) was more yield efficient, but less uniform efficient when compared with S2 lines recurrent selection (breeder's selection) for 'Pigarro'. In addition, phenotypic recurrent selection on 'Fandango', showed yield maximization during breeders selection (from cycle 1 to cycle 5) and a big ear size maximization by farmer's selection (after cycle 5). This indicates that farmers and breeders objectives/results are generally different. For this reason it is very important to set the criteria of selection at the beginning of a participatory plant breeding program.

2) Data from the evaluation trials of long term selection indicated phenotypic traits that better explain yield and identify a predictive model for yield. This would allow to reduce characterization costs, having the most representative traits. Traits that can help to predict yield based on maize ears;

3) Fasciation is present in the Portuguese maize traditional populations. Portuguese fasciation phenotypic studies existed, but its molecular basis was unknown. We identified several QTLs for traditional Portuguese maize ear fasciation in chromosomes 1, 3, 5, 7 and 8, with associated candidate genes *bif2*, *ra2*, *ra3*, *bd1* and *ts4*, *te1*, *bde1*, *ct1*, through linkage mapping. On the other hand, *dek19*, *dek28* and *mn3* were proposed as candidate genes for fasciation (Chapter V) through the 'Pigarro' genetic diversity evolution analysis. The linkage mapping analysis would ensure the use of a proper combination of genetic factors controlling ear diameter, kernel row number and ear length to allow ear fasciation expression without abnormal ear shapes and increase yield and/or ear size, depending on the final breeding objective. Newly detected QTLs represent interesting regions to further explore in maize yield research.

VIII.2. Context, challenges and future perspectives

VIII.2.1 Context

Portugal represents nearly 0.1% of the total production of maize in the world. Hence, Portugal has an important legacy in genetic

resources representing more than 500 years of coevolution adapted to human uses. The awareness of genetic erosion enhanced the collecting missions carried on since the 1970's in Portugal. The first missions were organized by Silas Pêgo. Pêgo was able to attract funds from FAO to build the first Portuguese germplasm bank and to provide funds to finance the venue of a genetic resources consulter (Rena Faria) (Pêgo 1996; Chapter I). At this time, also the need for germplasm improvement on-farm started to grow. With this purpose VASO project started in 1984. The VASO project allowed to improve germplasm (Amiúdo, Verdeal de Aperrela', 'Castro Verde', 'Pigarro' and 'Fandango') and create the link between farmers and breeders.

Along time researchers of ESAC-IPC, ITQB, INIAV and a farmers network were able to build the Portuguese Maize Cluster in which a multidisciplinary and transdisciplinary approach was established, making the convergence of targets and motivations, such as biodiversity, on-farm agroeco-systems, landscape, sustainable culture, polycropping, farming systems, quality aspects and human health (Belo et al. 2011; Belo 2012). The maize cluster works under Participatory Plant Breeding involving farmers, scientist, stakeholders and consumers, and promoting a multi-actor approach. The maize cluster activities have been possible due to national (FCT) and international funding (FP7 – SOLIBAM, H2020 - DIVERSIFOOD).

The Portuguese Maize Cluster action focus on the whole maize cycle from the environment (where the seed is sown) to the final product (*e.g.* maize bread) considering how can traditional landraces be kept

on farm and be improved without losing quality. Quality for maize bread is associated with taste and the structure of the final product. For this reason, it is important to have a feedback from consumers, which can happen, via participatory sensorial panels. Quality for maize bread also needed to have adequate food technology and the right raw material (Portuguese traditional maize landraces). These information's can be very important to define standards of quality to make the differentiation between maize for maize bread (*e.g.* traditional landraces) and maize for animal feeding (*e.g.* the majority of the hybrids). The adequate tools to monitor quality, can emphasize differentiation, promoting an adequate valorization of maize landraces. Landraces are also 'the living masterpieces' of the interaction among human, genotypes and environment representing traditions, its tastes and flavors (Negri 2005). These topics are also of great value for an adequate valorization. Which is needed, because there is a huge gap between landraces and modern cultivars' yield (*e.g.* maize) in most cases. This fact forces farmers to abandon their germplasm. Participatory plant breeding approaches (PPB) can be associated with *in situ* conservation of landraces contributing to their economically sustained presence in the farmers' fields. It can also contribute to define *in situ*/on-farm strategies that could help to design better synthetic hybrid populations for a new generation of low input and organic farming adapted to environmental changes and marginal areas.

VIII.2.2 Challenges and Future perspectives

To improve maize yield maintaining the quality some perspectives are indicated: the efforts to reduce stalk and root lodging should continue in a long term basis to insure acceptance of this germplasm by farmers. In farmers' selection particular attention should be given to maintain or reduce duration of cycles to avoid that yield could be improved at the cost of longer cycles with moisture increase, which increase drying costs. In addition, the research on the best traits selection for yield can be adapted to germplasm improvement by farmers under participatory plant breeding programs. In the majority of cases participatory plant breeding programs are associated with sustainable farming systems. These programs can enhance genetic resources (*e.g.* landraces) and respective genes combination for tolerance to pest and diseases and abiotic stresses, nutrient uptake efficiency. The phytonutrients and micronutrient concentrations generally present on the landraces indicate an adaptation to marginal conditions (*e.g.* protected areas) and to climate changes, due to its diversity and long-term adaptation representing a valuable potential in organic and low input farming (Maxted et al. 2002; Newton et al. 2010).

When farmers are involved in selection it is needed to enlighten the best traits to select for. These traits can eventually be important to predict yield when adequate formulas (*e.g.* ear value) are used. The yield prediction and respective formulas can be improved through instance ranking method. With this purpose it is important to test more diverse germplasm to increase data representativeness and

improve consistency of the studies already established. The study of heterotic groups among Portuguese germplasm or germplasm from other origins can be also of great important for future farmer's yield improvement, through hybrid populations' development. The development of hybrid populations could also contribute to yield progress and to avoid the loss of some germplasm. This approach can be applicable in a rural development strategy if economic benefits between associations for specialties (*e.g.* maize bread) and farmers could be achieved.

In addition, the plant density studies are also needed. This studies will help to adapt more appropriately the potential of a population to a certain environment, both *per se* or in a intercropping system. Furthermore the interaction with beneficial soil microorganisms' studies can be especially important in low input agriculture, improving plant nutrition.

Double haploid can help to obtain inbreeds for maize breeding programs on station but also can provide material for recombination if we chose a recurrent selection at farmers level maintaining diversity and promoting a dynamic population.

Future work should focus on the saturation of the genetic map here developed (non-fasciated PB260 x fasciated PB266), especially on the fasciated related QTL regions identified, with more codominant markers or other types of higher throughput dominant markers (such as Single Nucleotide Polymorphisms, SNP). This would allow to fill the gaps between distant markers, as well as increasing the likelihood of

merging the total amount of screened markers into 10 linkage groups. This would also increase the potential to identify possible candidate genes that can be used in Marker-assisted selection.

The limited inbred lines on this study indicate that some of the genes responsible for fasciation will not be represented. With this purpose it is suggested the use of Multiparent Advanced Generation Inter-Cross (MAGIC) (Cavanagh et al. 2008) where we can add new sources of fasciation (*e.g.* inbreeds of the Portuguese maize breeding program, double haploid lines of fasciated populations), but also to find the adequated combinations towards fasciation control. In this way fasciation can address commercial programs in a more easy way. Furthermore, fasciation through its adaptation to environment conditions can continue to be used in PPB programs.

Apart from the studies done it is very important that our needs match with the legislation available. Current intellectual property rights based on the "COMMISSION IMPLEMENTING DECISION of 18 March 2014 on the organization of a temporary experiment providing for certain derogations for the marketing of populations of the plant species wheat, barley, oats and maize pursuant to Council Directive 66/402/EEC' had open a time frame allowing to recognize the farmers' breeding efforts.

Yield and quality improvement of maize as many other species are slower processes that depend on a long term commitment to achieve the aimed results. When quality is the target and involves social aspects the complexity increases. This long term commitment

requires a cluster of farmers, scientists, millers, bakers, consumers, human health and others stakeholders where transdisciplinarity is part of the solution towards a renew interest on participatory plant breeding.

VIII.3. References

Allen SM, Jackson DP, Komatsu M, Pautler M, Sakai H, Vollbrecht E, Weeks R, Company EIDdNa, 2013. Nucleotide sequences encoding *fasciated ear*4 (fea4) and methods of use thereof. WO 2013138544 A1.

Belo M, Nobre A, Vaz Patto M, Boas L, da Silva S, Bronze MR (2011) Volatile and phenolic compounds in flour of maize varieties used in the production of traditional breads ("broa de milho"). In: EuroCereal 2011: Science and Technology Meeting Real World Challenges. Chipping Campden, Egyesült Királyság, 2011.12.06-2011.12.07.

Belo MMN (2012) Estudo de diferentes variedades de milho utilizadas na produção de broa. Tese de Mestrado da Faculdade de Farmácia da Universidade de Lisboa

Breiman L (2001) Random forests. Machine Learning 45: 5-32.

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and Regression Tree. Chapman and Hall/CRC.

Friedman JH (1991) Multivariate adaptive regression splines. Ann. Stat. 19: 1–141.

Cavanagh C, Morell M, Mackay I, Powell W (2008) From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants. Curr Opin Plant Biol 11: 215-221. doi: 10.1016/j.pbi.2008.01.002

Hallauer AR, Sears JH (1969) Mass Selection for Yield in Two Varieties of Maize1. Crop Sci 9: 47-50. doi: 10.2135/cropsci1969.0011183X000900010016x

Jackson DP, Nagasawa NS, Sakai H, Nagasawa N (2011) Nucleotide sequences encoding RAMOSA3 and sister of RAMOSA3 and methods of use for same. Google Patents.

Maxted N, Guarino L, Myer L, Chiwona E. 2002. Towards a methodology for on-farm conservation of plant genetic resources. Genet Resour Crop Ev 49: 31-46. doi: 10.1023/A: 1013896401710

Negri V (2005) Agro-Biodiversity Conservation in Europe: Ethical Issues. Journal of Agricultural and Environmental Ethics 18: 3-25.

Newton AC, Akar T, Baresel JP, Bebeli PJ, Bettencourt E, Bladenopoulos KV, Czembor JH, Fasoula DA, Katsiotis A, Koutis K, Koutsika-Sotiriou M, Kovacs G, Larsson H, Pinheiro de Carvalho MAA, Rubiales D, Russell J, Dos Santos TMM, Vaz Patto MC (2010) Cereal landraces for sustainable agriculture. A review. Agron. Sustain. Dev 30 (2): 237-269. doi: 10.1007/978-94-007-0394-0_10

Pautler M, Eveland AL, LaRue T, Yang F, Weeks R, Lunde C, Je BI, Meeley R, Komatsu M, Vollbrecht E, Sakai H, Jackson D (2015) FASCIATED EAR4 Encodes a bZIP Transcription Factor That Regulates Shoot Meristem Size in Maize. The Plant Cell Online. doi: 10.1105/tpc.114.132506

Pêgo SE (1982) Genetic potential of Portuguese maize with abnormal ear shape, Ph.D. Thesis, Iowa State Univ.

Pêgo SE (1996) Maize genetic resources in Portugal. In: Lipman E, Ellis RH, Gass T (eds), Maize genetic resources in Europe. Report of a workshop, Rome, Italy, IPGRI, pp 52-54.

Pêgo SE, Antunes MP (1997) Resistance or tolerance? Philosophy, may be the answer. In: Proceedings of the XIX – Conference of the International Working Group on Ostrinia. Guimarães Portugal 30th August–5th September 1997

Pêgo SE, Hallauer AR (1984) Portuguese maize germplasm with abnormal ear shape. Maydica 29: 39–53

Ribeiro de Brito JG (2014) Rankeamento de espigas de acordo com a produtividade estimada. Relatório de Mestrado Integrado em Engenharia Informática da Faculdade de Engenharia da Universidade do Porto

UPOV Test Guidelines http://www.upov.int/test_guidelines/en/list.jsp

Vaz Patto MC, Alves ML, Almeida NF, Santos C, Mendes-Moreira P, Šatović Z, Brites C. 2009. Is the bread making technological ability of portuguese traditional maize landraces associated with their genetic diversity? Maydica 54: 297-311. http://www.maydica.org/articles/54_297.pdf

Vaz Patto MC, Mendes-Moreira PM, Alves ML, Mecha E, Brites C, do Rosário Bronze M, Pêgo S. 2013. Participatory Plant Quality Breeding: An Ancient Art Revisited by Knowledge Sharing. The Portuguese Experience. In:

Andersen SB (ed) Plant Breeding from Laboratories to Fields, InTech. doi: 10.5772/52951

Vaz Patto MC, Moreira PM, Almeida N, Šatović Z, Pêgo S. 2008. Genetic diversity evolution through participatory maize breeding in Portugal. Euphytica 161:283-291. doi: 10.1007/s10681-007-9481-8