Exploring maize genetic diversity – a quality-driven perspective

Mara Lisa Alves

Dissertation presented to obtain the Ph.D degree in Biology

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

Oeiras, December, 2017

Work performed at:

Genetics and Genomics of Plant Complex Traits (PlantX) Laboratory

Instituto de Tecnologia Química e Biológica António Xavier

Universidade Nova de Lisboa

Av. da República, 2780-157 Oeiras, Portugal

Ph.D. Supervisor

Doctor Maria Carlota Vaz Patto

Head of the Genetics and Genomics of Plant Complex Traits Laboratory, Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal

Ph.D. Co-Supervisor

Professor Zlatko Šatović

Department of Seed Science and Technology, Faculty of Agriculture, University of Zagreb, Zagreb, Croatia

Table of contents

Acknowledgements

This thesis results from the cooperation established between many people coming from different backgrounds, and to whom I am very grateful. At times it was a challenge, but it was definitely a very fruitful one.

I would like to express my deep gratitude to Doctor Maria Carlota Vaz Patto, for accepting being my supervisor, for the unconditional support demonstrated over the years, for the input into our scientific discussions and by the patience shown even in the middle of my not-so-clear line of thought. I also would like to thank my co-supervisor, Professor Zlatko Šatović, for having accepted me as his student, for the training on genetic diversity analysis, and for receiving me so warmly during my stay at Zagreb University.

A word of appreciation to my thesis committee composed by Doctor Maria Manuela Veloso, Doctor Carla Brites, by my supervisor and co-supervisor, thank you for the support and for the scientific discussions along the way.

To all the other collaborators that allowed me to develop my Ph.D. work:

– From the Instituto Politécnico de Coimbra, Escola Superior Agrária de Coimbra: Doctor Pedro Mendes-Moreira, Cláudia Brites, Manuel Paulo, and Daniel Gaspar, for the establishment of the field trials, agronomic data acquisition, and for teaching me the basics on controlled pollinations in maize;

– From the Food Functionality and Bioactives Laboratory, ITQB NOVA: Professor Maria do Rosário Bronze, Maria Belo, Andreia Silva and Elsa Mecha, for quality data acquisition and for answering my endless questions;

– From Food Technology Laboratory, Instituto Nacional de Investigação Agrária e Veterinária: Doctor Carla Brites and Bruna Carbas, for quality data acquisition and for always being available;

– From the Biometris Laboratory, Wageningen University, The Netherlands: Professor Fred van Eeuwijk and Doctor Marcos Malosetti for the training on Genstat software, for the guidance in genome-wide association analysis, and for always being available to explain to a biologist the statistical slang;

– From the Department of Plant Breeding Genetics, Biometrics and Experimentation, Zagreb University, Croatia: Professor Jerko Gunjača, for the help in the multi-location field trials analysis.

A word of appreciation to Doctor Silas Pego and the farmers that were involved in the development of the maize populations evaluated in Chapter II and III, as well as to the Portuguese Bank of Plant Germplasm (Banco Português de Germoplasma Vegetal – BPGV, Braga) for providing some of the maize seeds used in the field trials of Chapters II, III, IV and V.

I would like to acknowledge all present and past members of the Genetics and Genomics of Plant Complex Traits Laboratory for the wonderful environment, friendship and support over the years. I would like to leave a special word of gratitude to Susana Leitão and Nuno Almeida, with whom I spent long hours “Bagin', bagin' you”.

I would like to acknowledge all present and past members of the Plant Cell Biotechnology Laboratory for the many cakes shared, and for the companionship over the years.

I would like to acknowledge Instituto de Tecnologia Química e Biológica António Xavier for the conditions, and for the scientific training provided under the ITQB Ph.D. Program.

To my “adoptive family”, Rui Malhó and Maria Leonor Reis, I would like to leave a word of gratitude for everything. Without

specifying anyone by fear of forgetting someone behind, I would like to acknowledge all the people in my personal circle of friends and family that have given me moral support throughout the years.

Thank you all!

List of most used abbreviations

% w/w	% mass per mass
µL	microliter
*a**	color red/green index
AL	total main volatile aldehydes (measured as the sum of the chromatogram peak area)
AM	*Amiúdo* maize population
AMMI	Additive Main effects and Multiplicative Interaction
AT	α-tocopherol
ATLOG	the logarithm of base 10 of the value for α-tocopherol
*b**	yellow/blue index (yellowness)
BD	breakdown viscosity (peak viscosity-trough viscosity)
BD_SqRt	Square-root transformation of the values for breakdown viscosity
BLUEs	best linear unbiased estimators
BLUPs	best linear unbiased predictors
bp	base pairs
BPGV	Portuguese Bank of Plant Germplasm
CA	*Castro Verde* maize population
chr	chromosome
cm	centimeter

cP	centiPoise
CU	*p*-coumaric acid
CW	cob weight
CWEW	cob weight/ear weight ratio
DT	δ-tocopherol
DTLOG	the logarithm of base 10 of the value for δ-tocopherol
E	ear placement
EM	ear moisture
EW	ear weight
FE	ferulic acid
FI	fiber
freq	frequency
FT	fat
FV	final viscosity
g	gram
G × E	genotype-by-environment interaction
mg GAE/100 g DW	milligrams of gallic acid equivalents per 100 grams of dry weight
GT	γ-tocopherol
GWAS	genome-wide association study
HPLC	high-performance liquid chromatography
HY	total hydroxycinnamic acids content
HYLOG	the logarithm of base 10 of the value for total

	hydroxycinnamic acids content
IPCA	Interaction Principal Component Analysis
kb	kilobase pairs
kg/ha	kilograms per hectare
*L**	lightness
LD	linkage disequilibrium
Ln	the natural logarithm
Log	the logarithm of base 10
m	meter
MAS	marker-assisted selection
Mb	megabase pairs
mg	milligram
mg FAE / 100 g DW	milligrams of ferulic acid equivalents per 100 grams of dry weight
Mg/ha	megagrams per hectare
min	minute
mL	milliliter
mL/min	milliliter per minute
mm	millimeter
N	leaf angle
nm	nanometer
ºC	degree Celsius
P	prolificacy

PB	participatory bred
PCA	principal component analysis
PCR	polymerase chain reaction
PH	total free phenolics
PHH	total free phenolics measured by high performace liquid chromatography
PHHLOG	the logarithm of base 10 of the value for total free phenolics measured by high performace liquid chromatography
PHS	total free phenolics measured with the *Folin–Ciocalteu* assay
PHSLOG	the logarithm of base 10 of the value for total free phenolics measured with the *Folin–Ciocalteu* assay
PPB	participatory plant breeding
PR	protein content
PV	peak (or maximum) viscosity
Q-Q	quantitle-quantile
QTL(s)	quantitative trait locus (loci)
R	root lodging
REML	REsidual Maximum Likelihood
RVA	Rapid Visco Analyser
S	stalk lodging
SB1	setback from trough (or minimum) viscosity
SB1	setback from trough viscosity (final viscosity -

	trough viscosity)
SB2	setback from peak viscosity (final viscosity - peak viscosity)
SIZE	mean particle size in non-lyophylized sample
SIZEL	mean particle size in lyophylized sample
SNP	single nucleotide polymorphism
SP	standing plants
SSR	simple sequence repeats or microsatellite
STL	starch content in lyophylized sample
ST	starch content in non-lyophylized sample
T	tassel branching
TCC	total carotenoid content
TV	trough (or minimum) viscosity
U	uniformity
US	United States of America
v/v	% volume per volume
VASO	Vale do Sousa-Sousa Valley (Portuguese participatory maize breeding program)
Y	grain yield
YP	grain yield per plant
µg/g	microgram per gram
µm	micrometer
σ^2	variance

Summary

In Portugal, and to a certain extent, farmers keep growing traditional maize populations known for their bread quality, conserving simultaneously their high genetic diversity. Maize populations, genetically more heterogeneous than commercial hybrid varieties, can evolve and better adapt to a changing broader range of edaphic-climatic conditions. Unfortunately, these maize populations suffer from a real risk of disappearing, due to their characteristic low yields. It is, therefore, desirable to improve their agronomic performance while maintaining their valuable diversity levels.

Important quality parameters, such as nutritional, organoleptic, and technological traits directly related to bread making ability, are generally characterized by a continuous variation and are highly influenced by the environment. This continuous variation suggests the influence of several genes, making them difficult to grasp by farmers and breeders.

The work settled in this thesis aimed to optimize selection approaches and develop molecular tools to assist on the implementation of participatory breeding programs focused on maize quality improvement, as a way to promote the on-farm conservation and improvement of the Portuguese maize populations.

To attain these objectives, the evolution of the genetic diversity during the improvement of two historical maize populations, Amiúdo and Castro Verde, was evaluated using microsatellites molecular markers. These populations have been subjected to on-farm stratified mass selection methodology for improving mainly agronomic traits, in the context of the Portuguese long-term participatory maize breeding program – VASO program. These molecular markers were also used to access the genetic diversity level present on other traditional maize

populations still under cultivation. These last populations were collected in the last decade from farmers located in a Portuguese region known to produce broa, a renowned maize-based bread. The molecular evaluation of all maize populations was further complemented by agronomic evaluations in multi-location field trials and by the evaluation of several quality-related parameters. Quality data - kernel color and composition (protein, fat, fiber), flour's pasting behavior, bioactive compound levels (carotenoids, tocopherols, phenolic compounds), and volatile aldehydes content - was assessed on flour of each population harvested from a common-garden experiment.

The results of the assessment of the effect of on-farm stratified mass selection in *Amiúdo* and *Castro Verde* populations revealed that this participatory program was able to improve or maintain populations' yield while preserving their genetic diversity. Nonetheless, it was also observed that the majority of the quality traits evaluated progressed erratically over selection time.

Agronomic, quality and molecular data allowed to evaluate the potential of traditional maize populations still under cultivation to be included in a quality-oriented participatory breeding program.

The quality characterization of Portuguese farmers' maize populations showed that these populations mainly presented high levels of protein and fiber content, low levels of carotenoids, volatile aldehydes, α- and δ-tocopherols content, and breakdown viscosity values. Regarding the agronomic performance, farmers' maize populations had low but considerably stable grain yields across the tested environments. As for their genetic diversity, each farmers' population was genetically heterogeneous. Nonetheless, all farmers' populations were molecularly distinct from each other's. The results from molecular, agronomic and quality evaluation were used to

generate a valuable tool to support an efficient and effective management of the available genetic resources in future breeding activities.

The difficulty to visually select for the majority of the quality traits considered in this work and the environmental influence in the resulting phenotype can be ameliorated by developing molecular markers linked to the trait of interest to support marker-assisted selection approaches. Through a genome-wide association study based on a collection of maize inbred lines partially derived from Portuguese maize populations, and using the phenotypic data obtained from 2 years of field trials and the genotypic information of 48,772 single nucleotide polymorphism markers from the MaizeSNP50 BeadChip array, it was possible to identify several genomic regions associated with quality-related traits. In the future, user-friendly molecular markers will be developed for the interesting genetic variants and these will be validated on a different genetic background in order for them to be useful for marker-assisted selection.

Concluding, the work developed under this Ph.D. thesis opened ways in the field of participatory maize breeding in Portugal, improved the knowledge on the quality characterization of traditional maize populations, postulating future paths for breeding these materials, and increased the basic and applied knowledge on the genetic control of quality-related traits in maize.

Sumário

Em Portugal, populações tradicionais de milho com grande qualidade para a produção de broa são ainda cultivadas por alguns agricultores, permitindo que a sua diversidade genética seja conservada. Estas populações de milho, geneticamente mais heterogéneas do que as variedades híbridas comerciais, adaptam-se mais facilmente a alterações edafo-climáticas. Encontram-se, contudo, em risco de desaparecerem devido aos seus baixos níveis de rendimento. Para reverter esta situação, é necessário melhorar o desempenho produtivo destas populações tradicionais sem comprometer os níveis de diversidade responsáveis pela resiliência geralmente associada a estes materiais.

As características de qualidade com influência direta na produção de broa, como por exemplo as características nutricionais, organoléticas e tecnológicas, apresentam geralmente uma variação contínua e são influenciadas por fatores ambientais. Essa variação contínua sugere que existem vários genes envolvidos responsáveis por essas características de qualidade, dificultando a tarefa de seleção a melhoradores e produtores.

Esta tese de doutoramento teve como objetivo otimizar os métodos de seleção e desenvolver ferramentas moleculares que facilitem a implementação de programas de melhoramento participativo direcionados para a melhoria da qualidade do milho, como forma de promover a conservação e desenvolvimento de populações portuguesas com melhor desempenho agronómico nos campos dos agricultores.

Para alcançar esses objetivos, foi primeiramente avaliada a evolução da diversidade genética de duas populações tradicionais portuguesas de milho, o Amiúdo e o Castro Verde, durante o

processo de melhoramento, utilizando marcadores moleculares do tipo microssatélite. Essas populações foram submetidas a uma metodologia de seleção massal estratificada, realizada *in-situ* com a participação de agricultores da região do Vale do Sousa, para melhoramento de características agronómicas, no contexto do programa português de melhoramento participativo a longo prazo – o programa VASO. Os marcadores moleculares utilizados permitiram também avaliar o nível de diversidade genética presente noutras populações tradicionais de milho ainda em cultivo, com potencial para serem incluídas num programa de melhoramento participativo orientado para a qualidade. Sementes destas populações foram obtidas, na década passada, contactando agricultores da região centro do país, região esta conhecida por produzir broa de elevada qualidade. A avaliação molecular destas populações de milho foi complementada por avaliações agronómicas em ensaios de campo multi-locais e pela avaliação de vários parâmetros relacionados com a qualidade realizada na farinha de cada população proveniente de um ensaio de campo - cor e composição do grão (proteína, gordura , fibra), características reológicas (viscosidade) da farinha, níveis de compostos bioativos (carotenoides, tocoferóis, compostos fenólicos) e conteúdo em aldeídos voláteis.

Os resultados da avaliação do efeito da seleção massal estratificada realizada *in-situ* nas populações Amiúdo e Castro Verde revelaram que este programa de melhoramento participativo foi capaz de melhorar ou manter o rendimento das populações, preservando sua diversidade genética. No entanto, também se observou que a maioria das características de qualidade avaliadas evoluiu erraticamente ao longo do tempo de seleção.

A caracterização da qualidade das populações de milho dos agricultores portugueses revelou que essas populações continham

altos níveis de proteína e de teor em fibra, e baixos níveis de carotenoides, aldeídos voláteis, α- e δ-tocoferóis, e baixos valores de viscosidade de degradação. Em relação ao desempenho agronómico, as populações de milho dos agricultores apresentaram baixos rendimentos de grão mas, no entanto, mais estáveis em todos os ambientes testados em comparação com as outras populações de milho analisadas. Quanto à sua diversidade genética, as populações de cada agricultor eram geneticamente heterogéneas. No entanto, todas as populações dos agricultores eram geneticamente distintas entre si. Os resultados das avaliações molecular, agronómica e de qualidade realizadas constituem uma ferramenta valiosa e fundamental para apoiar a conservação e gestão eficiente e efetiva dos recursos genéticos disponíveis em futuras atividades de melhoramento.

A dificuldade de selecionar visualmente a maioria das características de qualidade consideradas neste trabalho e a influência ambiental no fenótipo resultante podem ser ultrapassadas através do desenvolvimento de marcadores moleculares associados à característica de interesse, através de uma abordagem de seleção assistida por marcadores moleculares. Através de um estudo genético de associação realizado numa coleção de linhas puras de milho parcialmente derivadas de populações de milho portuguesas, utilizando os dados fenotípicos de 2 anos de ensaios de campo e a informação genotípica de 48.772 marcadores de polimorfismo de nucleotídeo único (SNPs) da plataforma de genotipagem MaizeSNP50 BeadChip*,* foi possível identificar várias regiões genómicas associadas às características de qualidade. No futuro, marcadores moleculares para as variantes genéticas de interesse, mais fáceis de usar por melhoradores, serão desenvolvidos e validados noutras populações de milho com diferente fundo genético,

para que sejam úteis para a seleção assistida por marcadores moleculares.

Concluindo, o trabalho desenvolvido durante esta tese de doutoramento abriu caminhos no campo do melhoramento participativo de milho em Portugal, aumentou o conhecimento sobre a caracterização da qualidade das variedades tradicionais portuguesas de milho, postulou caminhos futuros para o melhoramento desses recursos genéticos e contribuiu para o conhecimento básico e aplicado sobre o controlo genético de características relacionadas com a qualidade em milho.

Chapter I

General introduction

1 Quality in food crops – Considerations on maize

Maize (*Zea mays* L.) is, along with rice and wheat, one of the world's leading crops and a crucial source of food, feed, fuel, and fibers (Tenaillon & Charcosset, 2011). Maize is a staple for large populations in Latin America, Africa, and Asia (Ai & Jane, 2016), and the way maize kernels are processed and consumed varies greatly from country to country, with maize flour and meal being two of the most used products for producing many maize-based food commodities (Ai & Jane, 2016; Ranum et al., 2014).

When talking about food quality, several aspects can be considered and will all depend on the raw material composition and processing. This will ultimately depend on the end-use product. When using maize kernel for baking purposes, the improvement of the end-product quality can be achieved taking into consideration the upstream processes (e.g., genotype used, harvest procedures, seed quality, and pest control), but also the downstream processes (e.g., milling type, baking procedure). Although there are no clearly defined criteria for kernel quality for baking purposes (e.g., for maize bread), the kernel morphology and phytosanitary quality are generally considered as important (large grain size, uniformity, high density, and lack of physical damage, pests, and diseases) (Revilla et al., 2015, and references therein).

In some countries, such as Spain or Portugal, whole maize flour is used for bread production (Rodríguez et al., 2013). In Portugal, the ethnic maize-based bread is known locally as *broa*. *Broa* is traditionally made with more than 50% maize flour mixed with rye and/or wheat flour in a mostly empirical process (Brites et al., 2010). As further described by the same authors, this process normally involves the mixing of sieved wholemeal maize flour with hot water,

rye and/or wheat flour (in a variable proportion), with yeast from leavened dough from previous *broa* acting as sourdough.

In wheat, bread quality depends largely on the viscoelastic properties conferred to the dough by the gluten proteins (Shewry et al., 1995). However, maize has no gluten and the *broa* bread quality must be evaluated with different parameters. So, contrary to the wheat, maize dough has no viscoelastic network that enables to hold the gas produced during the fermentation process (Brites et al., 2010). Consequently, on maize flour, the parameters associated with bread quality cannot be evaluated as on wheat. On the absence or presence of a reduced amount of gluten, the dough rheological properties are provided by starch gelatinization (Brites et al., 2010).

Previously, Brites et al. (2010), through a sensory analysis on *broa* carried out by a trained panel using open-pollinated maize populations, identified a preference, due to texture, taste, and aroma, for maize bread produced using open-pollinated populations, as opposed to maize bread produced using commercial hybrid maize varieties. In the same study, instrumental quality attributes of maize flour from open-pollinated populations were measured and compared with commercial hybrid maize varieties. The results from that study showed that the flour from open-pollinated populations – considered by the trained panel to produce better quality *broa* – had higher values of protein, lower values of amylose, and lower viscosities (maximum, minimum, final, and breakdown viscosities) (Brites et al., 2010). More recently, several of the maize flour parameters that mainly influence maize quality for *broa* production have been identified (Carbas et al., 2016). Protein and amylose content, flour pasting parameters, such as maximum, minimum and final viscosities (Brites et al., 2010), but also flour particle size (Carbas et al., 2016),

are among these major influencing traits for the quality of the end-product.

1.1 The complex nature of maize kernel composition

The maize kernel is composed of four primary structures. They are endosperm, germ, pericarp, and tip cap make up 83%, 11%, 5%, and 1% of the maize kernel, respectively (Gwirtz & Garcia-Casal, 2014). Its high nutritional value is mainly due to its starch, protein, and oil content (Wen et al., 2016) but maize kernels are also rich in other micronutrients, such as vitamins (Nuss & Tanumihardjo, 2010).

Starch is maize's primary carbohydrate and kernel constituent (~72% of the kernel dry matter), consisting of a mixture of two polymers, amylose and amylopectin (reviewed by Nuss & Tanumihardjo, 2010). Protein is mostly distributed between the endosperm and the germ (~10% of the kernel dry matter; reviewed by Nuss & Tanumihardjo, 2010). Crude maize protein consists of a mixture of prolamins (called zeins), glutelins, albumins, and globulins, which are differentiated by their solubility properties. Prolamin is the major fraction, followed by glutelins, both of which are endosperm-specific proteins (reviewed by Nuss & Tanumihardjo, 2010). After starch and protein, fat in the form of oil is the third largest nutritional component of the kernel (~4% of the kernel dry matter) which is mainly concentrated in the germ (reviewed by Nuss & Tanumihardjo, 2010).

Crude fiber is highly present in the kernel seed coat (87% of the seed coat) but is also found in smaller amounts in the endosperm and germ walls (Nuss & Tanumihardjo, 2010). The majority of the maize fiber is dietary fiber, which is nearly completely insoluble (Rose et al., 2010). The insoluble dietary fiber in maize is mainly composed of cellulose and hemicellulose, with only a small amount of lignin (Rose

et al., 2010). In recent years, dietary fiber has attracted increasing interest, as many studies have revealed that it may be involved in disease preventing and health promoting activities (reviewed in Sun et al., 2015).

Additionally, micronutrients such as vitamins, are found in all major parts of the kernel, including the endosperm (provitamin A carotenoids), germ (vitamin E), and aleurone (water-soluble vitamins) (reviewed by Suri & Tanumihardjo, 2016). Vitamin A, as provitamin A carotenoids, and vitamin E, as tocopherols, are the predominant fat-soluble vitamins found in maize kernels. Both carotenoids and tocopherols play important roles as antioxidants (reviewed by Nuss & Tanumihardjo, 2010). Even though carotenoids are yellow-orange phytopigments, kernel color is not necessarily correlated with provitamin A concentration in orange and yellow cultivars, due to its variable accumulation in the seed coat, endosperm, and germ (Harjes et al., 2008). As for vitamin E, it is found almost exclusively in the maize germ oil at about 94% of total tocopherols (reviewed by Nuss & Tanumihardjo, 2010). For most varieties, α- and γ-tocopherols are the only vitamin E constituents found in significant amounts (reviewed by Nuss & Tanumihardjo, 2010).

1.2 How kernel composition affects processing

Pasting properties of maize flour are considered important parameters in the preparation of different food products as they are related to its swelling and gelatinization ability (Paraginski et al., 2014). Starch, proteins, and lipids are the three major food components in cereal-based food products, and interactions among them in a food system are of importance to functionality and quality (Wang et al., 2017; Zhang & Hamaker, 2003).

Changes in starch biochemical characteristics, such as the proportion and structure of amylose and amylopectin, will influence its viscosity and gelatinization ability, determining the kernel uses in distinct products such as bread, beer, or biopolymers (reviewed by Cozzolino, 2016). Fiber content can also have an impact on baked goods quality, contributing to dough viscosity, air entrapment and the improvement of loaf volume and texture (Rose et al., 2010). As reviewed by Cozzolino (2016), the presence or addition of chemicals can also modify starch properties. For example, the texture and structural stability of starch-based raw materials can be modified due to interactions between starch with fatty acids (reviewed by Cozzolino, 2016). Also the presence of antioxidant phenolic compounds may alter and improve starch qualities (Beta & Corke, 2004; Siriamornpun et al., 2016; Zhu et al., 2009), or influence the dough texture (Klepacka & Fornal, 2006), a very important parameter in defining bread quality (Matos & Rosell, 2012).

In numerous maize-based foods, the endosperm or kernel hardness has been described as having a major impact on quality (Carbas et al., 2016; de la Hera et al., 2013; Fox & Manley, 2009). The size of the particles that are released from flour is directly related to the kernel hardness. Harder kernels or those richer in vitreous endosperm will yield larger particles than those that are softer (Chandrashekar & Mazhar, 1999). With regard to the biochemical contribution to maize kernel hardness, both protein and starch composition are implicated, and specifically, the variation in zein classes has been linked to differences in hardness (reviewed in Fox & Manley, 2009). The content and composition of zeins are the key determinants of protein quality and texture-related traits of the kernel (Wang et al., 2016).

2 Molecular breeding for maize quality

Maize breeding has primarily focused on increasing stability and grain yield potential, under abiotic and biotic stresses (reviewed in Muzhingi et al., 2017). In the last decade, however, much effort has been made in evaluating and using the diversity of maize also on the improvement of animal feed and human nutrition (reviewed in Muzhingi et al., 2017). Currently, maize breeding efforts for improved chemical composition is being extended beyond the traditional targets of starch, oil, and protein to include components such as vitamins, and antioxidant secondary metabolites with considerable consequences for human health (Wen et al., 2016). By using marker-assisted selection, a few nutritional trait-associated genes or QTLs (for maize protein quality, oil content and provitamin A levels) have been recently introgressed into elite maize lines for their quality improvement (Wen et al., 2016, Table 2 therein).

As reviewed by Moose and Mumm (2008), conventional plant breeding that relies only on phenotypic selection has been historically effective on crop improvement. However, for some traits, phenotypic selection has made little progress due to challenges in phenotype accurate measurement or in the identification of the individuals with the highest breeding value. The effects of environment and genotype-by-environment interaction also contribute to the reduced progress in conventional plant breeding. For some traits, only destructive measurements are available to accurately access the phenotype, or trait expression may be dependent on the developmental stage (e.g., kernel quality traits) (Moose & Mumm, 2008).

Bread quality parameters are generally characterized by a continuous variation, suggesting the influence of several genes. It is, thus, expected that several of the *broa*'s quality parameters show

quantitative inheritance. Quantitative traits cannot be classified into discrete phenotypic classes, making it impossible to use Mendelian approaches. The identification and location of genes controlling these traits through Quantitative Trait Loci (QTL) analysis can overcome this difficulty (Prioul et al., 1997). The genetic architecture of complex quantitative traits is generally studied with the final objective of improving crop performance (Yang et al., 2010). Functional markers are developed and applied in molecular breeding programs (through marker-assisted selection) after the identification of favorable alleles by linkage analysis or association mapping (Andersen & Lübberstedt, 2003).

QTL linkage mapping approaches suffer from two fundamental limitations. First, only the allelic diversity that segregates between the parents of a particular cross can be assayed, and second, the amount of recombination that occurs during the development of linkage mapping populations places a limit on the mapping resolution (reviewed in Korte & Farlow 2013). In genome-wide association studies, the rapid breakdown of linkage disequilibrium among diverse maize lines (association panel) is exploited, enabling very high resolution for QTL mapping via association analysis (Flint-Garcia et al., 2005). In maize, several QTL linkage mapping studies, and for the last 15 years, association mapping studies, were successfully undertaken on nutritional quality and have shown that kernel main components and other health-related compounds (e.g., tocopherols and carotenoids) are controlled by many genes, having complex patterns of inheritance (e.g., Cook et al., 2012; Diepenbrock et al., 2017; Jittham et al., 2017; Li et al., 2013). The elucidation of the genes underlying flour main components variation is essential for efficiently improving this crop quality.

3 Maize in Portugal – a long and diverse story

Maize, a naturally open-pollinated species, was one of the first crops to be domesticated more than 9,000 years ago in the valleys of Mexico (Matsuoka et al., 2002). After domestication, maize spread rapidly across the Americas (Mir et al., 2013). The first historical record attesting to the introduction of maize to Europe is dated from 1493 when Columbus brought it from the Caribbean to Spain (Tenaillon & Charcosset, 2011). Also according to historical records, this crop rapidly reached other European countries such as Italy during the 15th century (Brandolini & Brandolini, 2009), and Portugal by the beginning of the 16th century (Oliveira, 1999). Once maize production was established, centuries of evolution in small farm households gave rise to a variety of landraces, or traditional maize populations across the country. As reviewed in Vaz Patto et al. (2013), each traditional maize population can be defined as an open-pollinated population with an associated historical origin and a distinct identity, lacking any formal crop improvement, as well as often being genetically diverse, locally adapted, and associated with traditional farming systems. According to genetic studies, the Portuguese traditional maize populations seemed to be the result of multiple introduction events from at least two distinct geographic origins, consisting nowadays of a mixture of material from the Caribbean islands and material from northeastern America (e.g., Rebourg et al., 2003; Dubreuil et al., 2006).

After World War II, Portugal was one of the first European countries to test the US maize hybrids which initially were not well accepted by the Portuguese farmers due to several handicaps such as late maturity or kernel type, not fitted for food or polycropping systems (Vaz Patto et al., 2013). Nevertheless, several breeding

stations were established within Portugal at that time, from North to South, in the cities of Braga (NUcleo de melhoramento de MIlho - NUMI), Porto, Viseu, Elvas and Tavira, that soon started to release adapted hybrid varieties based on inbreds developed from Portuguese and US germplasm (reviewed in Vaz Patto et al., 2013). Seeds from the maize germplasm developed within those breeding stations are currently curated by the Portuguese Bank of Plant Germplasm (Banco Português de Germoplasma Vegetal - BPGV, Braga, Portugal).

After the 1986, when Portugal joined the European Union, changes in the agriculture policy — the introduction of monocropping systems, the valorization of crop uniformity and yield, the high mechanization and fossil inputs, with low manpower, the increase land parcel area, with a close market-oriented output for feed — led to a replacement of the traditional maize populations by hybrid varieties. This replacement by hybrid varieties put the Portuguese maize landraces in real risk of disappearance. Fortunately, part of this germplasm was already conserved at the BPGV, through an initiative during the 70's that aimed to collect locally grown traditional maize populations. Additionally, some enduring landraces also survived at the farmers' fields due to particular traits (Vaz Patto et al., 2007). Farmers traditionally select maize seed based on their intrinsic bread quality, ear size, and aspect, yield, pests and lodging resistance, and maintain a high level of variability to ensure yield under any conditions (Vaz Patto et al., 2013).

To provide an incentive for *in-situ* conservation of traditional maize landraces, Silas Pego, at the beginning of the 80s, had the idea of engaging local farmers and their seeds in a maize participatory plant breeding program (PPB). By doing this, his goals were not only to conserve but also to improve the social well-being of

this rural community by increasing farmers' income through rising yields using their own seeds. To bring that idea to practice he led, in 1984, a detailed survey on farmer's maize fields at Sousa Valley Region, in the Northwest of Portugal. The collected materials were the starting point of a PPB project, with simultaneous on-farm breeding and on-farm conservation objectives (VASO - ''Vale do Sousa''- project). This project aimed 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 US intensive agriculture model did not fit and the seed multinationals had no adequate market to operate (Vaz Patto et al., 2013).

Nowadays, with the development of modern sustainable low-input agriculture in industrialized countries, for economic and environmental reasons, an emphasis has been placed on local adaptation, on the preservation of genetic diversity, and on quality (Cleveland et al., 1999). Conventional plant breeding has been successful in favorable environments, but is less successful in traditional low-input or organic farming systems with higher stress growing conditions, especially in small-scale farms (Vaz Patto et al., 2013). Under this scenario, participatory plant breeding (PPB) programs are arising worldwide to meet the needs of farmers in low-input and organic environments that are normally overlooked by conventional crop breeders (Vaz Patto et al., 2013).

Participatory plant breeding differs from conventional breeding mainly because of the active participation of other actors apart from breeders, such as farmers and/or consumers, in the breeding program. Those actors can assume an active role in the establishment of the breeding objectives and influence or actively participate in the breeding activities. In the case of on-farm

participatory breeding, the selection is made at the farmer's field, in a partnership between breeder and farmer, with the farmer establishing the breeding objectives (Vaz Patto et al., 2013). This type of decentralized PPB improves breeding efficiency as it increases the ratio of the number of varieties adopted by farmers, as it is the farmer's choice to adopt those varieties into the program; it also increases traits' response to selection, as selection is being made in the targeted environment (Ceccarelli, 2015).

In the specific case of the Portuguese maize PPB program, the impact of breeding activities on the maize populations' agronomic performance improvement has until now only been measured in two out of the several maize populations in the program (Mendes-Moreira et al., 2008, 2009), and the temporal changes in genetic diversity were only evaluated for one of those populations (Vaz Patto et al., 2008). Moreover, none of these studies took into consideration quality aspects that should be addressed in future breeding programs since the quality of these genetic resources for maize bread production seems to be a decisive aspect for the on-farm maintenance of the historical populations developed and for their present market added-value (Brites et al., 2010; Vaz Patto et al., 2013).

In the 21st century, Portuguese traditional maize populations can be still found under production as verified in a collecting expedition that took place in the last decade in the Central-Northern region of Portugal (Vaz Patto et al., 2007). This mission had as main objective sampling the enduring traditional maize populations' variability of a particular region in the country, where maize-based bread still plays an important role in the local rural economy. Most of the traditional Portuguese maize landraces are white flints and potentially with a good technological ability for *broa* bread production (Vaz Patto et al., 2007). In this collecting expedition, the majority of

the maize populations conserved were being used primarily for bread production. Around 50 different (yellow/orange and white) maize landraces were collected, characterized using pre-breeding approaches and conserved in cold storage (Vaz Patto et al., 2007). This collection was later enlarged with landraces collected subsequently from the surrounding regions. The fact that the flour produced from locally grown maize populations has traditionally been used in the formulation of *broa* has been pointed out by Vaz Patto et al. (2007) as one of the reasons for the on-farm conservation of the Portuguese maize populations. As a consequence, the collected populations were assumed to have the potential to be used in *broa* production. Several other features of traditional maize populations have been identified by other authors to explain why these populations are still maintained under cultivation, such as the fact that when compared to hybrids, maize landraces have a broader plasticity to adapt to different environments (Hellin et al., 2014).

The endured Portuguese collected maize landraces represent important sources of genes and gene combinations not yet available for crop quality breeding programs. These materials due to their intrinsic quality traits (that promoted their maintenance in cultivation) are the best candidates for expanding the already existing participatory breeding program (VASO) to other regions with particular emphasis on quality breeding. However, little is known about the phytochemical profiles, antioxidant activity, or organoleptic quality of the different Portuguese maize open-pollinated populations with a high technological ability for bread production.

Besides the phenotypic characterization, a better understanding of the genetic diversity present in the germplasm available for breeding will help to structure germplasm, defining, for example, heterotic pools. In addition, it will provide useful information for

selecting contrasting parental lines for new breeding populations and will help breeders to identify valuable new alleles for breeding (Varshney et al., 2016). For an effective conservation and management of these interesting plant resources and an effective use in quality breeding programs, it is fundamental to: understand the parameters affecting bread quality; study the genetic control/basis of these complex traits visually difficult to select; and characterize the Portuguese collected maize landraces diversity.

4 Objectives and outline of the present study

This thesis is built upon the evaluation of nutritional quality (macro and micronutrients) and processing traits directly or indirectly related with *broa* bread quality, that is dependent on the composition of the wholemeal maize flour, and the used of molecular information to build decision-making tools directed towards the establishment of a quality-based participatory breeding program. The definition of this objective was partly supported by several related aspects (1) the unique and diverse maize germplasm existence in the country, conserved both *in-situ* and *ex-situ*; (2) the empirical knowledge that *broa* made from Portuguese open-pollinated varieties have distinct quality characteristics not present in *broa* from modern commercial varieties; (3) the analytical and instrumental knowledge on several measurable, physicochemical parameters that distinguish and influence the maize bread produced with these maize populations; (4) and the availability of high-throughput genotyping platforms developed for maize.

This Ph.D. thesis focused on the phenotypic and genotypic characterization of a variety of maize germplasm (populations and inbred lines) to allow the development of molecular-based tools for breeding purposes. It is restricted to a set of quality-related traits that

were previously identified as being particularly important for bread quality in maize. These traits fall mainly in one of these two categories: nutritional quality (macro and micronutrients) and processing quality traits, both measured at the wholemeal flour level, since it is known that the former (flour quality) will influence the later (bread quality).

With this thesis, molecular tools, together with phenotypic data (agronomic and quality) were used to estimate the effect of on-farm stratified mass selection on the agronomic performance, quality, and molecular diversity of two historical maize populations; and to characterize the genetic diversity of Portuguese maize landraces. A maize inbred line collection partially derived from Portuguese maize landraces was also used to perform a whole-genome association study to identify genomic regions/candidate genes associated with traits related to maize bread quality.

Therefore, the specific objectives of this work were:

1) To evaluate if on-farm stratified mass selection, in the context of long-term participatory research, was able to improve the agronomic performance of two historical maize open-pollinated populations, Amiúdo and Castro Verde;

(2) To evaluate the effect of stratified mass selection in the genetic diversity levels of these two populations;

(3) To evaluate the effect of stratified mass selection in quality traits (related to consumer preferences, technological, nutritional, and organoleptic properties) that may influence maize bread quality.

(4) To extend the maize populations quality characterization – organoleptic, nutritional, and health-related traits – with the quantification of aroma-related volatile compounds, and health-

related compounds, such as tocopherols, carotenoids, and phenolic compounds, that might influence the quality and the consumer acceptability of maize-based food commodities;

(5) To accurately estimate the agronomic performance and potential of the collected enduring maize populations using multi-location field trials (broader performance stability/specific adaptability) across different farming sites, exploring new locations for the establishment of a future quality-oriented participatory maize breeding program;

(6) To build decision-making tools to enable an accurate population selection within a quality-oriented participatory breeding program, based on the integration of agronomic, quality and molecular characterization of the maize populations;

(7) To identify genomic regions controlling for quality-related traits through a genome-wide association approach; and

(8) To identify putative candidate genes involved in each trait variation.

The thesis is structured as follows: in Chapter II the results on the temporal genetic stability of two maize populations under long-term stratified mass selection are presented — this information was also integrated with the evolution on quality and agronomic performance of those maize populations bred under an on-farm participatory breeding program. Chapter III highlights how the integration of agronomic, quality and molecular data can potentially be used as a decision-making tool in a future quality-oriented participatory maize breeding program. Chapters IV and V present the identification of genomic regions controlling for nutritional and technological traits (Chapter IV), and for health-related (antioxidant) compounds (Chapter V) in wholemeal maize flour. Finally, in Chapter

VI, the thesis main achievements, key lessons, and action points for future work are identified and discussed.

References

Ai, Y., & Jane, J. L. (2016). Macronutrients in corn and human nutrition. *Comprehensive Reviews in Food Science and Food Safety*, *15*(3), 581-598.

Andersen, J. R., & Lübberstedt, T. (2003). Functional markers in plants. *Trends in Plant Science*, *8*(11), 554-560.

Beta, T., & Corke, H. (2004). Effect of ferulic acid and catechin on sorghum and maize starch pasting properties. *Cereal Chemistry*, 81(3), 418-422.

Brandolini, A., & Brandolini, A. (2009). Maize introduction, evolution and diffusion in Italy. *Maydica*, 54(2), 233-242.

Brites, C., Trigo, M. J., Santos, C., Collar, C., & Rosell, C. M. (2010). Maize-based gluten-free bread: influence of processing parameters on sensory and instrumental quality. *Food and Bioprocess Technology*, *3*(5), 707-715.

Carbas, B., Vaz Patto, M. C., Bronze, M. R., Bento da Silva, A., Trigo, M. J., & Brites, C. (2016). Maize flour parameters that are related to the consumer perceived quality of 'broa' specialty bread. *Food Science and Technology*, 36, 259–267.

Ceccarelli, S. (2015). Efficiency of plant breeding. *Crop Science*, 55, 87–97.

Chandrashekar, A., & Mazhar, H. (1999). The biochemical basis and implications of grain strength in sorghum and maize. *Journal of Cereal Science*, 30(3), 193-207.

Cleveland, D. A., Soleri, D., & Smith, S. E. (1999). Farmer plant breeding from biological perspective: Implications for collaborative plant breeding. CIMMYT Economics Working Paper CIMMYT, Mexico DF: CIMMYT, 99–10.

Cook, J. P., McMullen, M. D., Holland, J. B., Tian, F., Bradbury, P., Ross-Ibarra, J., et al. (2012). Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. *Plant Physiology*, 158(2), 824-834.

Cozzolino, D. (2016). The use of the rapid visco analyser (RVA) in breeding and selection of cereals. *Journal of Cereal Science*, *70*, 282-290.

de la Hera, E., Talegón, M., Caballero, P., & Gómez, M. (2013). Influence of maize flour particle size on gluten-free breadmaking. *Journal of the Science of Food and Agriculture*, *93*(4), 924-932.

Diepenbrock, C. H., Kandianis, C. B., Lipka, A. E., Magallanes-Lundback, M., Vaillancourt, B., Góngora-Castillo, E., et al. (2017). Novel Loci Underlie Natural Variation in Vitamin E Levels in Maize Grain. *The Plant Cell* Online, tpc-00475.

Dubreuil, P., Warburton, M., Chastanet, M., Hoisington, D., & Charcosset, A. (2006). More on the introduction of temperate maize into Europe: large-scale bulk SSR genotyping and new historical elements. *Maydica*, 51, 281-291.

Flint-Garcia, S. A., Thuillet, A. C., Yu, J., Pressoir, G., Romero, S. M., Mitchell, S. E., et al. (2005). Maize association population: a high-resolution platform for quantitative trait locus dissection. The Plant Journal, 44(6), 1054-1064.

Fox, G., & Manley, M. (2009). Hardness methods for testing maize kernels. *Journal of Agricultural and Food Chemistry*, 57(13), 5647-5657.

Gwirtz, J. A., & Garcia-Casal, M. N. (2014). Processing maize flour and corn meal food products. *Annals of the New York Academy of Sciences*, *1312*(1), 66-75.

Harjes, C. E., Rocheford, T. R., Bai, L., Brutnell, T. P., Kandianis, C. B., Sowinski, S. G., et al. (2008). Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science*, *319*(5861), 330-333.

Hellin, J., Bellon, M. R., & Hearne, S. J. (2014). Maize landraces and adaptation to climate change in Mexico. *Journal of Crop Improvement*, 28, 484–501.

Jittham, O., Fu, X., Xu, J., Chander, S., Li, J., & Yang, X. (2017). Genetic dissection of carotenoids in maize kernels using high-density single nucleotide polymorphism markers in a recombinant inbred line population. *The Crop Journal*, 5(1), 63-72.

Klepacka, J., & Fornal, Ł. (2006). Ferulic acid and its position among the phenolic compounds of wheat. *Critical Reviews in Food Science and Nutrition*, 46(8), 639-647.

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, 9(1), 29.

Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., et al. (2013). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nature genetics*, 45(1), 43-50.

Liu, R. H. (2003). Health benefits of fruit and vegetables are from additive and synergistic combinations of phytochemicals. *The American Journal of Clinical Nutrition*, *78*(3), 517S-520S.

Matos, M.E., & Rosell, C.M. (2012). Relationship between instrumental parameters and sensory characteristics in gluten-free breads. *European Food Research and Technology*, 235(1), 107-117.

Matsuoka, Y., Vigouroux, Y., Goodman, M.M., Sanchez, J., Buckler, E., & Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences*, 99(9), 6080-6084.

Mendes-Moreira, P. M. R., Pego, S. E., Vaz Patto, C., & Hallauer, A. R. (2008). Comparison of selection methods on 'Pigarro', a Portuguese improved maize population with fasciation expression. *Euphytica*, 163, 481–499.

Mendes-Moreira, P. M. R., Vaz Patto, M. C., Mota, M., Mendes-Moreira, J., Santos, J. P. N., Santos, J. P. P., et al. (2009). 'Fandango': Long term adaptation of exotic germplasm to a Portuguese on-farm-conservation and breeding project. *Maydica*, 54, 269–285.

Mir, C., Zerjal, T., Combes, V., Dumas, F., Madur, D., Bedoya, C., et al. (2013). Out of America: tracing the genetic footprints of the global diffusion of maize. *Theoretical and Applied Genetics*, 126(11), 2671-2682.

Moose, S. P., & Mumm, R. H. (2008). Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiology*, 147(3), 969-977.

Muzhingi, T., Palacios-Rojas, N., Miranda, A., Cabrera, M. L., Yeum, K. J., & Tang, G. (2017). Genetic variation of carotenoids, vitamin E and phenolic compounds in Provitamin A biofortified maize. *Journal of the Science of Food and Agriculture*, 97(3), 793-801.

Nuss, E. T., & Tanumihardjo, S. A. (2010). Maize: a paramount staple crop in the context of global nutrition. *Comprehensive Reviews in Food Science and Food Safety*, *9*(4), 417-436.

Oliveira, A. (1999). Nos rumos da modernidade (Vol. 4). Universidade do Porto, Portugal.

Paraginski, R. T., Vanier, N. L., Berrios, J. D. J., De Oliveira, M., & Elias, M. C. (2014). Physicochemical and pasting properties of maize as affected by storage temperature. *Journal of Stored Products Research*, 59, 209-214.

Prioul, J. L., Quarrie, S., Causse, M., & de Vienne, D. (1997). Dissecting complex physiological functions through the use of molecular quantitative genetics. *Journal of Experimental Botany*, *48*(6), 1151-1163.

Ranum, P., Peña-Rosas, J. P., & Garcia-Casal, M. N. (2014). Global maize production, utilization, and consumption. *Annals of the New York Academy of Sciences*, 1312(1), 105-112.

Rebourg, C., Chastanet, M., Gouesnard, B., Welcker, C., Dubreuil, P., & Charcosset, A. (2003). Maize introduction into Europe: the history reviewed in the light of molecular data. *Theoretical and Applied Genetics*, 106(5), 895-903.

Revilla, P., de Galarreta, J. I. R., Malvar, R. A., Landa, A., & Ordás, A. (2015). Breeding maize for traditional and organic agriculture. *Euphytica*, 205(1), 219-230.

Rodríguez, V. M., Soengas, P., Landa, A., Ordás, A., & Revilla, P. (2013). Effects of selection for color intensity on antioxidant capacity in maize (*Zea mays* L.). *Euphytica*, 193(3), 339-345.

Rose, D. J., Inglett, G. E., & Liu, S. X. (2010). Utilisation of corn (*Zea mays*) bran and corn fiber in the production of food components. *Journal of the Science of Food and Agriculture*, 90(6), 915-924.

Siriamornpun, S., Tangkhawanit, E., & N. Kaewseejan. (2016). Reducing retrogradation and lipid oxidation of normal and glutinous rice flours by adding mango peel powder. *Food Chemistry*, 201, 160-167.

Siró, I., Kápolna, E., Kápolna, B., & Lugasi, A. (2008). Functional food. Product development, marketing and consumer acceptance—A review. *Appetite*, 51(3), 456-467.

Sun, Q., Wu, M., Bu, X., & Xiong, L. (2015). Effect of the amount and particle size of wheat fiber on the physicochemical properties and gel morphology of starches. *PloS one*, 10(6), e0128665.

Shewry, P. R., Tatham, A. S., Barro, F., Barcelo, P., & Lazzeri, P. (1995). Biotechnology of breadmaking: unraveling and manipulating the multi-protein gluten complex. *Nature Biotechnology*, *13*(11), 1185-1190.

Suri, D. J., & Tanumihardjo, S. A. (2016). Effects of different processing methods on the micronutrient and phytochemical contents of maize: from A to Z. *Comprehensive Reviews in Food Science and Food Safety*, *15*(5), 912-926.

Varshney, R.K., Singh, V.K., Hickey, J.M., Xun, X., Marshall, D.F., Wang, J., et al. (2016). Analytical and decision support tools for genomics-assisted breeding. *Trends in Plant Science*, 21(4), 354-363.

Vaz Patto, M.C., Mendes-Moreira, P.M., Alves, M.L., Mecha, E., Brites, C., Bronze, R., et al. (2013). Participatory plant quality breeding: An ancient art revisited by knowledge sharing. The Portuguese experience. In *Plant breeding from laboratories to fields,* S. B. Andersen (Ed.), (pp. 255–288). Rijeka,Croatia: InTech.

Vaz Patto, M. C., Moreira, P. M., Almeida, N., Satovic, Z., & Pego, S. (2008). Genetic diversity evolution through participatory maize breeding in Portugal. *Euphytica*, 161, 283–291.

Vaz Patto, M. C., Moreira, P. M., Carvalho, V., & Pego, S. (2007). Collecting maize (*Zea mays* L. convar. mays) with potential technological ability for bread making in Portugal. *Genetic Resources and Crop Evolution*, *54*(7), 1555-1563.

Wang, G., Wang, G., Wang, J., Du, Y., Yao, D., Shuai, B., et al. (2016). Comprehensive proteomic analysis of developing protein bodies in maize (*Zea mays*) endosperm provides novel insights into its biogenesis. *Journal of Experimental Botany*, erw396.

Wang, S., Zheng, M., Yu, J., Wang, S., & Copeland, L. (2017). Insights into the Formation and Structures of Starch–Protein–Lipid Complexes. *Journal of Agricultural and Food Chemistry*, 65(9), 1960-1966.

Wen, W., Brotman, Y., Willmitzer, L., Yan, J., & Fernie, A. R. (2016). Broadening our portfolio in the genetic improvement of maize chemical composition. *Trends in Genetics*, *32*(8), 459-469.

Tenaillon, M.I., & Charcosset, A. (2011). A European perspective on maize history. *Comptes Rendus Biologies*, 334(3), 221-228.

Yang, X., Yan, J., Shah, T., Warburton, M. L., Li, Q., Li, L., et al. (2010). Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. *Theoretical and Applied Genetics*, *121*(3), 417-431.

Zhang, G., & Hamaker, B. R. (2003). A three component interaction among starch, protein, and free fatty acids revealed by pasting profiles. *Journal of Agricultural and Food Chemistry*, 51(9), 2797-2800.

Zhu, F., Cai, Y.Z., Sun, M., & Corke, H. (2009). Effect of phytochemical extracts on the pasting, thermal, and gelling properties of wheat starch. *Food Chemistry*, 112(4), 919-923.

Chapter II

Temporal genetic stability of two maize (*Zea mays* L.) populations under long-term stratified mass selection

The work presented in this chapter was published in the following research publication:

Alves, M. L., Belo, M., Carbas, B., Brites, C., Paulo, M., Mendes-Moreira, P., Brites, C., Bronze, M. R., Šatović Z., & Vaz Patto, M. C. (2017). Long-term on-farm participatory maize breeding by stratified mass selection retains molecular diversity while improving agronomic performance. *Evolutionary Applications* 00:1–17. doi: 10.1111/eva.12549

In this research paper, Mara Lisa Alves performed the DNA isolation and the simple sequence repeat markers genotyping, the molecular diversity analysis, the statistical data analysis, and drafted the manuscript. (See Acknowledgements section for authors' contributions)

Abstract

Modern maize breeding programs gave rise to genetically uniform varieties that can affect maize's capacity to cope with increasing climate unpredictability. Maize populations, genetically more heterogeneous, can evolve and better adapt to a broader range of edaphic–climatic conditions. These populations usually suffer from low yields; it is therefore desirable to improve their agronomic performance while maintaining their valuable diversity levels. With this objective, a long-term participatory breeding/on-farm conservation program was established in Portugal. In this program, maize populations were subject to stratified mass selection. This work aimed to estimate the effect of on-farm stratified mass selection on the agronomic performance, quality, and molecular diversity of two historical maize populations. Multi-location field trials, comparing the initial populations with the derived selection cycles, showed that this selection methodology led to agronomic improvement for one of the populations. The molecular diversity analysis, using microsatellites, revealed that overall genetic diversity in both populations was maintained throughout the selection. The comparison of quality parameters between the initial populations and the derived selection cycles was made using the kernel from a common-garden experiment. This analysis showed that the majority of the quality traits evaluated progressed erratically over time. In conclusion, this breeding approach, through simple and low-cost methodologies, proved to be an alternative strategy for genetic resources' on-farm conservation.

Keywords: ear traits, microsatellites, molecular diversity, on-farm conservation, open-pollinated populations, participatory plant breeding, yield, *Zea mays* L.

1 Introduction

Climate change represents a challenge to food security (Wheeler & von Braun, 2013). The negative impact of climate change on agriculture and therefore on food production is exacerbated by greater crop uniformity (Ceccarelli et al., 2010). An increasing number of studies show that biodiversity improves the capacity of agroecosystems to cope with extreme weather events and climate variability (Khoury et al., 2014; Ortiz, 2011), allowing crops' evolution and adaptation to specific edaphic–climatic conditions (Ceccarelli, 2015). This is particularly important in the context of low-input/ organic production systems, more prone to biotic and abiotic constraints and in which crop resilience is fundamental. The greater uniformity of crops is specifically a concern for maize, wheat, and rice, which alone provide 60% of the calories in the human diet. In these three crops, recent plant breeding has led to extreme genetic uniformity (Ceccarelli et al., 2013). As reviewed by Hellin et al. (2014), it is important that plant breeding reach a compromise by developing not only higher-yielding but also stress-tolerant cultivars, to allow them to cope and adapt when faced with different environmental conditions. In the case of maize, the more heterogeneous open-pollinated populations, adapted to specific environmental conditions and human uses, have progressively been replaced in the last century by homogeneous, higher-yielding commercial hybrids (Pingali, 2001). Still, open-pollinated populations' cultivation has been maintained, often in marginal lands or low-input systems where commercial hybrids are not well adapted (Vaz Patto et al., 2013). They may also be kept by their dietary or nutritional value, taste, or for the price premium they attract because of high-quality traditional properties that compensate for lower yields (Jarvis et al., 2011).

Portugal was one of the first European countries to adopt maize and one of the few where historical maize populations can still be found under cultivation (Vaz Patto et al., 2013). The resilience of these maize populations in the Portuguese scenario can be partially explained by their technological quality in maize bread production (Vaz Patto et al., 2013). The Portuguese ethnic maize-based bread, named *broa*, is highly accepted for its distinctive sensory characteristics (Carbas et al., 2016). This bread is traditionally manufactured using local maize populations and still plays an important economic and social role in Central and Northern rural communities of the country (Vaz Patto et al., 2007). *Broa* is traditionally made with more than 50% maize flour mixed with rye and/or wheat flour by a mainly empirical process (Brites et al., 2010). This process normally involves the mixing of the sieved wholemeal maize flour, with hot water, rye and/or wheat flour (in a variable proportion), and yeast from leavened dough from late *broa*, acting as sourdough (Brites et al., 2010).

In what concerns *broa* bread quality, differences between the higher-yielding dent hybrids and the hard endosperm Portuguese open-pollinated populations have been recently determined (Carbas et al., 2016). In that work, it was shown that the *broa* produced with the hybrid dent varieties had higher specific volume. However, a sensory analysis showed a preference for the maize bread made using Portuguese open-pollinated populations due to better mouthfeel flavor and texture (Carbas et al., 2016). Parameters associated with aroma or flavor (e.g., volatile aldehydes; Klensporf & Jelén, 2005), and texture (e.g., viscosity parameters; Brites et al., 2010) can be important in assessing the product's quality and therefore need to be investigated. Additionally, bread nutritional value is another quality aspect of great importance. In recent years, consumption of particular

foods and food products, rich in antioxidant compounds, has been associated with the prevention of modern lifestyle-related degenerative disease (Liu, 2003). In that regard, maize displays a considerable natural variation in the content and composition of antioxidant compounds such as carotenoids (Owens et al., 2014) and tocopherols (Lipka et al., 2013). However, little is known about the phytochemical profiles, antioxidant activity, or organoleptic quality of the different Portuguese maize open-pollinated populations with a high technological ability for bread production.

With the development of modern sustainable low-input agriculture in industrialized countries, for economic and environmental reasons, an emphasis has been placed on local adaptation, on the preservation of genetic diversity, and on quality (Cleveland et al., 1999). Conventional plant breeding has been successful in favorable environments, but is less successful in traditional low-input or organic farming systems with higher stress growing conditions, especially in small-scale farms (Vaz Patto et al., 2013). Under this scenario, participatory plant breeding (PPB) programs are arising worldwide to meet the needs of farmers in low-input and organic environments that are normally overlooked by conventional crop breeders (Vaz Patto et al., 2013).

Participatory plant breeding differs from conventional breeding mainly because of the active participation of other actors apart from breeders, such as farmers and/or consumers, in the breeding program. Those actors can assume an active role in the establishment of the breeding objectives and influence or actively participate in the breeding activities. In the case of on-farm participatory breeding, the selection is made at the farmer's field, in a partnership between breeder and farmer, with the farmer establishing the breeding objectives (Vaz Patto et al., 2013). Taking into

consideration the central role attributed to farmers on this breeding approach, their acceptance, and enthusiasm while participating in the program has been identified as one of the key aspects for the success of on-farm participatory plant breeding (Vaz Patto et al., 2013). This type of decentralized PPB improves breeding efficiency as it increases the ratio of the number of varieties adopted by farmers, as it is the farmer's choice to adopt those varieties into the program; it also increases traits' response to selection, as selection is being made in the targeted environment (Ceccarelli, 2015).

In 2012, Ceccarelli et al. (2012) published the results of a survey on the previous major PPB experiences worldwide. Of the 22 active PPB programs presented in that report, three are in maize and are located in Portugal, China, and Nepal. The Portuguese participatory maize breeding program started in 1984 and initially had as its main objective the improvement of the agronomic performance of historical maize populations, functioning in parallel as a strategy for the on-farm conservation of those plant genetic resources (Vaz Patto et al., 2013).

The methodologies implemented in every breeding program are dependent on the type of reproductive system of the crop. In naturally cross-pollinated species, such as maize, improvement of open-pollinated populations can be achieved by recurrent mass selection if the pollinations are controlled and/or by the use of stratified selection (Gardner, 1961). In the on-farm breeding activities of the Portuguese maize participatory breeding program, as controlled pollinations are time-consuming, the use of stratified mass selection has been the selected methodology. In mass selection, a fraction of individuals is visually selected to form the following generation. As for stratified mass selection, prior to the selection of individuals (mass selection), the field is first divided into smaller selection units (field stratification),

minimizing the bias due to field heterogeneity. The differences among plants within field's sections are more likely to be due to genetic differences than to environmental effects (Hallauer et al., 2010). Stratified mass selection has been shown in the past to be a useful methodology for improving several agronomic traits in maize, for example, for adapting exotic germplasm into breeding programs and target environments (Hallauer, 1999) or for yield improvement of open-pollinated maize populations (Mendes-Moreira et al., 2008, 2009; Smith et al., 2001).

In the Portuguese maize participatory breeding program, breeding activities were intended to occur mainly in the farmer's field, with breeder and farmer working side by side. Firstly, the selection methodologies were demonstrated by the breeder at each farmer's field, and afterward, the farmer conducted the same selection methodologies in the other part of the field. In this way, the farmer had a permanent possibility to compare the effectiveness of the breeder's advice and the breeder needed to respect the farmer's management system (e.g., low-input), advising only simple and low-cost selection methodologies based on population genetics theory, with the farmer keeping the decision power over the direction of selection. Besides the specific breeding objectives defined by each farmer for each maize population, in this program the farmer is advised by the breeder to select in the field by detasseling the undesirable plants before pollination (weakest and all that do not fit the desired ideotype, such as the pest and disease susceptible looking ones); the farmer is also advised to evaluate a few days before harvest the root and stalk quality by foot-kicking the plants at their base (at the first visible internodes). This also serves as an indirect measurement of pest tolerance, as the plant that does not resist the impact and breaks down is eliminated. Additionally, the

farmer is advised to favor the selection of more prolific plants or the ones with a lower ear insertion if that trait is among the farmer desired ideotype. Prior to this selection, the field is first divided into smaller selection units (field stratification). After harvesting, a second selection (postharvest) is conducted in the ears. This selection includes the specific breeding objectives of each population and the elimination of unhealthy damaged ears. Selected ears are then shelled and mixed together to form the next-year generation. With this scheme, the selection pressure ranges from 1% to 5% (Mendes-Moreira et al., 2009). Generally, the postharvest selection is the only selection that the farmer traditionally carries out (nonformal selection) and the one that had been applied to the historical maize populations previously to their introduction in this participatory program.

2 Materials and Methods

2.1 Populations' origin and main features

The two historical open-pollinated maize populations evaluated in this study were previously subjected to on-farm stratified mass selection in the context of a participatory breeding program. This breeding program has been running in Portugal since 1984 in the Sousa Valley region, in the northern part of the country. Each maize population in this breeding program occupied, on average, an area of 1,000 m^2 and was composed of approximately 5,000 individuals per growing season (given a plant density of 50,000 plants/ha).

Amiúdo, a yellow flint early population (FAO 200), was chosen to integrate the PPB program in its beginning, in 1984. This population was selected due to its short life cycle and because it had already adapted to the local conditions (poor soils with low pH, water

stress, and aluminum toxicity); it was also chosen because it could be used for bread production (Vaz Patto et al., 2013).

The Amiúdo population was selected at two different locations: at the Lousada site (41°14′7.8″N 8°18′11.1″W), where the selection was performed by the breeder and farmer; and at the Serra do Carvalho site (41°34′12.74″N, 8°19′28.77″W), where the selection was performed by the breeder. In both cases, the specific breeding objective, set by the farmer, was to achieve a higher-yielding population; the same selection methodologies were applied at both the Lousada and Serra do Carvalho sites.

Castro Verde, an orange flint late population (FAO 600), was introduced in the PPB program in 1994 with the initial aim of achieving a population that could run in the category of yellow flint in a contest for the “Best Ears” of the Sousa Valley. This population was characterized by its big ears and very tall plants (>3 m in height).

Until 2000, Castro Verde was selected at the Lousada site (41°14′7.8″N 8°18′11.1″W) by the farmer. The selection criteria were set to obtain bigger ears by improving the traits that might enable the ears to win the “Best Ears” contest, namely ear length and kernel weight, row number, and the number of kernels per ear. After 2001, due to a reduction in the breeding activities at the Lousada site, the Castro Verde population began to be selected at the Coimbra site (40°13′0.22″N, 8°26′47.69″W) by the breeder. At that point, some adjustments were made to the breeding objectives but keeping the same selection methodologies (stratified mass selection). Specifically, selection criteria were fine-tuned to decrease the height of the ear insertion on the stalk, increase the stalk resistance, and keep increasing the ear size while still maintaining an orange flint kernel.

As a result of 19 years of Amiúdo selection at Lousada site, 19 cycles of stratified mass selection were originated, and as a result of

25 years of Amiúdo selection at Serra do Carvalho site, 25 cycles of stratified mass selection were originated. In this study, the following Amiúdo cycles were analyzed: the initial population from 1984, considered as cycle 0 (hereafter referred to as $AM_{C0\text{-}1984}$), and the nineteenth and the twenty-fifth cycles of stratified mass selection, obtained in 2003 at the Lousada site (hereafter referred to as AM-$L_{C19\text{-}2003}$) and in 2009 at the Serra do Carvalho site (hereafter referred to as AM-$SC_{C25\text{-}2009}$), respectively.

As a result of 14 years of Castro Verde selection, 14 cycles of stratified mass selection were originated between Lousada and Coimbra sites. In this study, the following Castro Verde cycles were analyzed: the initial population from 1994, considered as cycle 0 (hereafter referred to as $CA_{C0\text{-}1994}$), and the ninth and fourteenth cycles of stratified mass selection at Coimbra obtained in 2004 (hereafter referred to as CA-$C_{C09\text{-}2004}$) and in 2009 (hereafter referred to as CA-$C_{C14\text{-}2009}$), respectively.

The summary of the specific breeding objectives for the Amiúdo and Castro Verde populations, as well as the timeline and selection sites where the different cycles, analyzed in this work, were developed, is given in Figure 1.

2.2 Agronomic evaluation

The agronomic performance of two historical maize populations, Amiúdo and Castro Verde, and their derived selection cycles was compared in multi-location field trials. The Amiúdo initial population ($AM_{C0\text{-}1984}$) and selection cycles (AM-$L_{C19\text{-}2003}$ and AM-$SC_{C25\text{-}2009}$) were evaluated in eight locations: Quinta da Conraria, Montemor-o-Velho, S. Pedro do Sul, Lousada, Valada do Ribatejo, Vouzela-1, Vouzela-2, and Travassos. The Castro Verde initial population ($CA_{C0\text{-}1994}$) and selection cycles (CA-$C_{C09\text{-}2004}$ and CA-$C_{C14\text{-}2009}$) were evaluated in five

locations: Quinta da Conraria, Montemor-o-Velho, Lousada, Valada do Ribatejo, and Covas do Monte.

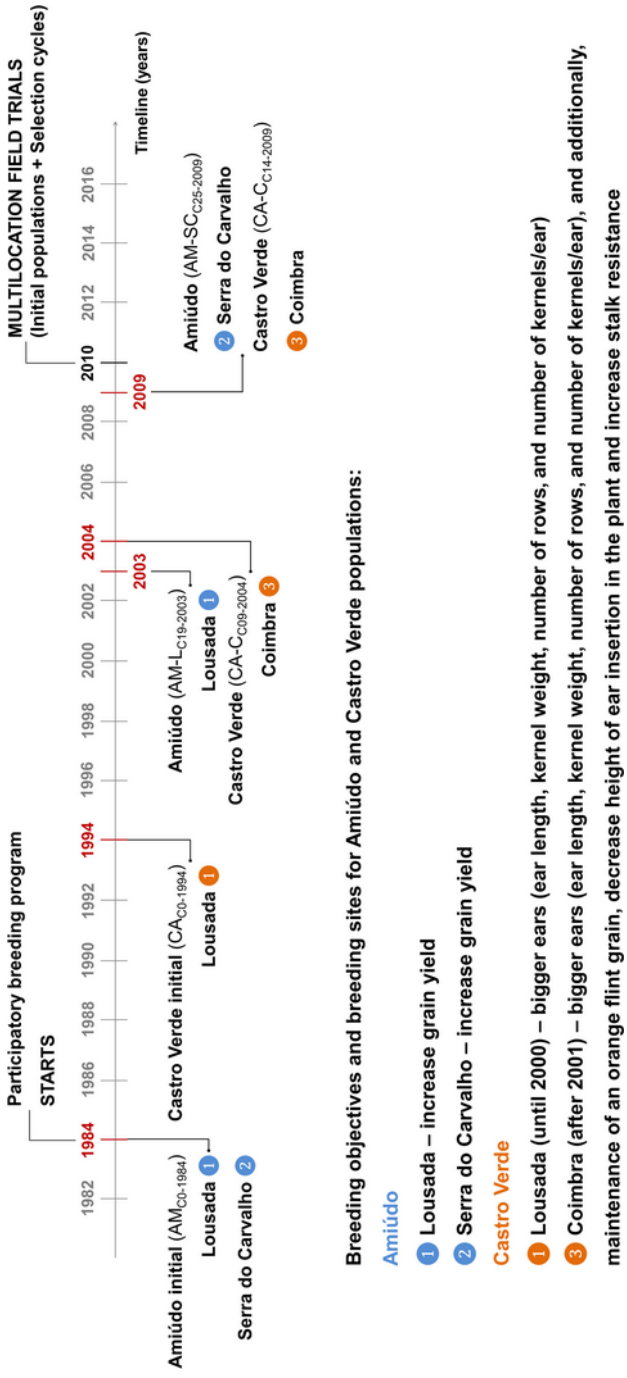

Figure 1. Breeding objectives, timeline, and selection sites for the analyzed Amiúdo cycles (initial population – $AM_{C0\text{-}1984}$; $AM\text{-}L_{C19\text{-}2003}$ selection cycle; and $AM\text{-}SC_{C25\text{-}2009}$ selection cycle) and Castro Verde cycles (initial population – $CA_{C0\text{-}1994}$; $CA\text{-}C_{C09\text{-}2004}$ selection cycle; and $CA\text{-}C_{C14\text{-}2009}$ selection cycle).

The different locations represent different areas where maize open-pollinated populations are traditionally produced in the country and also the different agronomic production systems normally associated with maize open-pollinated populations, ranging from conventional production systems (Montemor-o-Velho) to organic production systems (Quinta da Conraria and Valada do Ribatejo) to low-input production systems (all the other locations). Information about the sites' characterization is given in Table S1. Initial populations and selection cycles were evaluated, at farmers' fields, in a randomized complete block design, with three blocks per location. Each initial population and derived selection cycles were overplanted by hand in two-row plots 6.4 m long and with 0.75 m between rows. Each plot was thinned at the seven-leaf stage to 48 plants per plot to achieve a plant density of 50,000 plants/ha. Therefore, in each environment, a total of 144 plants (48 plants per plot*3 blocks) were evaluated for each cycle. Plots were irrigated as needed and mechanically weeded and/or hand-weeded as necessary. All the plots were harvested by hand.

The agronomic evaluation of each initial population and derived selection cycles was performed as described in Table 1. The data collected were intended to track eventual changes occurring in ear morphology, plant architecture, plant health and quality of the stalk and root system, population uniformity, and grain production.

Table 1. List of agronomic traits evaluated per plot basis, codes, and respective description.

Type of trait	Trait	Code	Units/Scale	Description
Ear morphology	Ear weight	EW	Gram (g)	Ear weight, adjusted to 15% of grain moisture. Measure by averaging the weight of 4 shelled ears per plot.
	Cob weight	CW	Gram (g)	Cob weight, adjusted to 15% of grain moisture. Measured by averaging the weight of the cobs of 4 shelled ears per plot.
	Cob weight/ear weight	CWEW	Ratio (g/g)	Ratio cob/ear weight indicates the proportion of cob weight in the ear weight. This ratio was taken from the cob and ear weights of 4 shelled ears per plot.
	Ear moisture	EM	Percentage (%)	Measured with a FARMPOINT moisture meter, using a mixture sample of 4 shelled ears grain per plot.
Plant architecture	Ear placement	E	1-9 scale	Ear placement in the plant. In this scale a 5 indicates that the 1st ear is located in the middle of the plant; values < 5 indicates that 1st the ear is located bellow the plant middle point; values > 5 indicates that the 1st ear is located above the plant middle point. This value was measured by evaluating all plants per plot.
	Leaf angle	N	1-9 scale	Angle of the adaxial side of the leaf above the ear with the stalk. In this scale a 5 indicates a leaf angle = 45 °; values < 5 indicate a leaf angle <45 °; and values > 5 indicate a leaf angle > 45 °. This value was measured by evaluating all plants per plot.
	Tassel branching	T	1-9 scale	In this scale 1 indicates unbranched tassel (typical of inbred lines) and 9 indicates a highly branched tassel (frequent in populations with fasciated ears). This value was measured by evaluating all plants per plot.

Continuation of Table 1

Type of trait	Trait	Code	Units/Scale	Description
Health and quality of the stalk and root system	Root lodging	R	Percentage (%)	Root lodging corresponds to percentage of plants leaning more than 30° from vertical in each plot. This value was measured by evaluating all plants per plot.
	Stalk lodging	S	Percentage (%)	Stalk lodging corresponds to percentage of plants broken at or below the primary ear node. This value was measured by evaluating all plants per plot.
	Standing plants	SP	No. plants/hectare (no. plants/ha)	Estimation of the number of standing plants per hectare given the number of plants at harvest time in the area of each plot (9.6 m^2).
Population uniformity	Uniformity	U	1-9 scale	Measure of population uniformity. In this scale 1 indicates minimum uniformity and 9 indicates maximum uniformity. Values from 1 to 4 are typical of open-pollinated populations and values from 5 to 9 are typical of pure lines. Measured by evaluating all plants per plot.
Grain production	Prolificacy	P	No. ears/plant	Total number of ears per plot divided by the total number of plants per plot.
	Grain yield	Y	Kilogram/hectare (kg/ha)	Grain yield adjusted to 15% moisture. Formula: Grain yield = Ear weight × (Grain weight/Ear weight) × (100%–% moisture at harvest)/ (100%–15% moisture). Grain weight and ear weight taken from 4 shelled ears.
	Grain yield per plant	YP	Gram/plant (g/plant)	Grain yield adjusted to 15% moisture divided by the number of standing plants per hectare.

2.3 Agronomic data analysis

All agronomic data analysis was carried out in SAS software (SAS Release 9.2.; SAS Institute, 2004).

Analysis of variance for Amiúdo cycles (initial population—$AM_{C0\text{-}1984}$; AM-$L_{C19\text{-}2003}$ selection cycle; and AM-$SC_{C25\text{-}2009}$ selection cycle) and for Castro Verde cycles (initial population—$CA_{C0\text{-}1994}$; CA-$C_{C09\text{-}2004}$ selection cycle; and CA-$C_{C14\text{-}2009}$ selection cycle) was carried out separately per population using the PROC MIXED procedure. In the mixed-model statement, environments and cycles (initial population and derived selection cycles) were treated as fixed effects, while blocks, treated as random, were nested in the environments. The interaction between cycles and the environment was included in the model. Cycle means were compared using a Tukey–Kramer multiple comparisons test.

To summarize multivariate changes occurring in both populations across the participatory breeding program, a principal component analysis (PCA) on the standardized agronomic data was performed using the PROC PRINCOMP procedure. The number of principal components was determined by inspecting eigenvalues of principal components (using the Kaiser criterion that retains components with eigenvalues greater than one). The first two principal components were then projected in a biplot to display shifts occurring in the agronomic traits measured on both initial populations and their selection cycles.

2.4 Molecular evaluation

Thirty random individual plants from the Amiúdo and Castro Verde initial populations and derived selection cycles were genotyped with 20 microsatellites (SSRs—simple sequence repeats). SSRs were chosen based on their location in the maize reference genome

(1 SSR per chromosome arm) and repeat motifs (≥3 base pairs) to facilitate allele scoring (Table S2). Information about each SSR can be found at MaizeGDB (Lawrence et al., 2008, www.maizegdb.org).

DNA was isolated from adult leaves of each plant using the modified CTAB procedure as described in Saghai-Maroof et al. (1984). DNA quality was accessed using a 0.8% SeaKem® LE Agarose gel (Cambrex Bio Science Rockland, Inc., USA) stained with SYBR® Safe (Invitrogen, USA). DNA quantification was performed using a spectrophotometer, Nanodrop ND-2000C (Thermo Scientific, USA). An additional step for polysaccharide removal (Rether et al., 1993) was added when the ratio 260/230 nm wavelength was inferior to 1.6 to avoid the interference of these contaminants in SSR amplification.

The SSR loci were amplified using a nested-PCR method (Schuelke, 2000). PCR products were separated on 6.5% polyacrylamide sequencing gel (20 µl 6.5% KB^{Plus} Gel Matrix, 150 µl APS 10%, and 15 µl TEMED) using a LI-COR 4300 DNA analyzer system. To account for any variance between PCR amplifications and electrophoresis runs, DNA from the B73 maize inbred line was used as a reference sample. Scoring of the alleles was confirmed manually by two independent users to ensure scoring accuracy. A genotypic matrix of the alleles per individual plant, scored in base pairs, was generated and served as the basis for the molecular data analysis.

2.5 Molecular data analysis

To assess the intracycle genetic diversity, the average number of alleles per locus (N_{av}), observed (H_O) and expected heterozygosity (H_E), and inbreeding coefficient (F_{IS}) were calculated for each initial population and selection cycles using GENEPOP software (GENEPOP v4.0; Raymond & Rousset, 1995). The values of these

estimates, obtained in each initial population and selection cycles, were then compared to test whether the values of N_{av}, H_O, H_E, and F_{IS} were significantly different among cycles with the Kruskal–Wallis test using SAS software (SAS Release 9.2, SAS Institute Inc 2004).

The genotypic frequencies for each locus and for each Amiúdo and Castro Verde cycles were tested for conformance to Hardy–Weinberg (HW) expectations using GENEPOP software (GENEPOP v4.0; Raymond & Rousset, 1995). The probability test was based on the Markov chain method (Guo & Thompson, 1992; Raymond & Rousset, 1995) using 10,000 dememorization steps, 20 batches, and 5,000 iterations per batch. The sequential Bonferroni adjustments (Rice, 1989) were then applied to correct for the effect of multiple tests using SAS software (SAS Release 9.2, SAS Institute Inc 2004).

Differences in allele frequencies distributions along the breeding program were tested according to Waples (1989a), in which the null hypothesis states that the observed differences in allele frequency can be explained entirely by genetic drift and sampling error. For the Amiúdo population, the temporal variation in allele frequencies was tested (i) between the Amiúdo initial population ($AM_{C0\text{-}1984}$) and the selection cycle from the Lousada site ($AM\text{-}L_{C19\text{-}2003}$), and (ii) between the Amiúdo initial population ($AM_{C0\text{-}1984}$) and the selection cycle from the Serra do Carvalho site ($AM\text{-}SC_{C25\text{-}2009}$). For the Castro Verde population, the temporal variation in allele frequencies was tested between the initial Castro Verde population ($AM_{C0\text{-}1984}$) and the latter selection cycle from the Coimbra site ($CA\text{-}C_{C14\text{-}2009}$). Afterward, the sequential Bonferroni adjustments (Rice, 1989) were applied to the level of significance to correct for the effect of multiple tests using SAS software (SAS Release 9.2, SAS Institute Inc 2004). The effective population size, which is a parameter necessary to test for temporal variation in allele frequencies,

according to Waples (1989a), was estimated using NeEstimator software (NeEstimator v2.01, Do et al., 2014) following the temporal-based method under sample plan II (Waples, 1989b), as the samples analyzed did not return to the breeding program. Alleles with a frequency lower than 0.05 were excluded, parametric chi-squared 95% confidence intervals for effective population size were calculated, and the variance in allele frequencies was calculated according to Nei and Tajima (1981).

Analysis of molecular variance (AMOVA; Excoffier et al., 1992), a method of estimating population differentiation directly from molecular data, was used to test whether the different cycles from Amiúdo and Castro Verde populations had suffered genetic differentiation along the breeding program. This was done by testing the partition of the total microsatellite diversity between and within each pair of cycles, as well as among and within all cycles using ARLEQUIN software (ARLEQUIN v3.0; Excoffier et al., 2005). The variance components retrieved from AMOVA were used to calculate a series of statistics called ϕ-statistics, which summarize the degree of differentiation between population divisions and are analogous to Wright's *F*-statistics (Excoffier et al., 1992). The variance components were tested statistically by nonparametric randomization tests using 10,000 permutations in ARLEQUIN software (ARLEQUIN v3.0, Excoffier et al., 2005).

To represent genetic relationships among individual plants, a factorial correspondence analysis (FCA) was carried out using GENETIX software (GENETIX v4.05; Belkhir et al., 2004), as this analysis provides a way of visually showing how genetically distant the different initial populations and derived selection cycles are; it also serves as a method for observing the level of genetic homogeneity within each cycle.

2.6 Quality evaluation

As both populations are used for human consumption, we also measured in each of the Amiúdo and Castro Verde initial populations and derived selection cycles several traits associated with kernel quality. Therefore, this study also intended to evaluate in which way traits related to flour's pasting behavior (flour viscosity parameters), nutritional value (protein, fat, and fiber content), potential bioactive compounds (carotenoids, tocopherols, total phenolic compounds content), and aroma-related compounds (volatile aldehydes) have changed or were maintained along the PPB program. For that, a bulk of kernel from each selection cycle produced from a common-garden experiment established in Coimbra in 2009, under controlled pollinations, was used.

Wholemeal maize flour was obtained after milling the kernel through a Falling number 3100 mill (Perten, Sweden), using a 0.8-mm screen.

2.6.1 Pasting behavior

The pasting properties of maize flour were obtained with a Rapid Viscosity Analyzer RVA-4 (Newport Scientific, Australia) at 15% solids as described in Brites et al. (2010). Peak (PV), minimum or trough (TV), and final viscosities (FV) were recorded in cPoise, and the breakdown (BD) was calculated as PV-TV.

2.6.2 Flour color parameters

Maize flour color was determined on 10–12 g of sample in an opaque recipient using a Minolta chromameter CR-2b and CIE tristimulus color parameters: *L**—lightness; *a**—red/green index; and *b**—yellow/blue index. *L** values can vary from *L** = 0 (black) to *L** = 100 (white); positive *a** values mean that samples tend toward the red

part of the color spectra; positive *b** values mean that samples tend toward the yellow part of the color spectra.

2.6.3 Protein, fat, and fiber content

Flour protein (PR), fat (FT), and fiber (FI) content were determined by a near-infrared spectroscopic method with an Inframatic 8620 equipment (Perten, Sweden), with calibrations supplied by the manufacturer. Results were expressed in percentage.

2.6.4 Total carotenoid content

The total carotenoid content (TCC) was spectrophotometrically measured at 450 nm according to the AACCI method 14-60.01 (AACC International 2012). Results were expressed in μg of lutein equivalent per gram of sample, as the main carotenoid found in maize.

2.6.5 Tocopherols content

α-Tocopherol (AT), γ-tocopherol (GT), and δ-tocopherol (DT) were separated from the fat portion of the maize flours by high-performance liquid chromatography (HPLC) and quantified using an Agilent 1200 model with a fluorescence detector (FLD) and a Diol column (LiChropher 100, 250 × 4 mm) according to the method ISO 9936 (2006). Tocopherols content was expressed in μg/g fat basis.

2.6.6 Total free phenolic content

Ethanolic extracts (EtOH:H_2O 50:50, v/v) for assessing the total phenolic content (PH) of maize flour were prepared as described in Lopez-Martinez et al. (2009), with some modifications as described in detail in Supplementary Material.

The total free phenolic content was assessed using the Folin–Ciocalteu assay (Singleton et al., 1999) with a Beckman DU-70

spectrophotometer, with slight modifications as described in Silva et al. (2015), and expressed in mg of gallic acid equivalents/100 g of dry weight (GAE/100 g DW).

2.6.7 p-*Coumaric and ferulic acid content*

p-Coumaric (CU) and ferulic acid (FE) were quantified by HPLC coupled with a photodiode array detector (HPLC-PDA) at 280 nm with a Thermo Finnigan Surveyor HPLC system according to Silva et al. (2006). *p*-Coumaric (CU) and ferulic acid contents were expressed in mg/100 g of dry weight (mg/100 g DW).

2.6.8 Volatile aldehydes content

The volatile fraction of maize flour was analyzed by solid-phase microextraction–gas chromatography–mass spectrometry (SPME-GC-MS). A 2-cm 50/30-μm DVB/Carboxen/PDMS fiber (SUPELCO) was used for solid-phase microextraction. Volatile compounds were analyzed with a GCMS-QP2010 Plus Shimadzu equipment and separated in a Varian Factor Four column (30 m × 0.25 mm × 0.25 μm). Volatile aldehydes content (AL) was taken as the sum of the peak area of the main aldehydes identified (hexanal, heptenal, 2-heptanal (*Z*), 2-octenal (*E*), nonanal, 2-nonenal (*E*), and decanal). Details on the quantification of volatile aldehydes content can be found in Supplementary Material.

2.7 Quality data analysis

To summarize the eventual multivariate changes on the evaluated quality traits occurring in both populations across the participatory breeding program, a principal component analysis (PCA) was performed using the PROC PRINCOMP procedure after standardization of the quality traits, similar to what has been already described for the agronomic data analysis.

3 Results

In this work, the agronomical, molecular, and quality evolution of two historical open-pollinated maize populations, Amiúdo and Castro Verde, across a participatory plant breeding program was accessed.

3.1 Agronomic evolution

In relation to the Amiúdo population agronomic performance, on-farm stratified mass selection led, in both selection sites—Lousada and Serra do Carvalho—to a significant increase in ear (EW) and cob weight (CW) and cob/ear weight ratio (CWEW) (0.9%–1.2% for EW, 2.1%–3% for CW, and 1%–1.6% gain per cycle for CWEW, respectively) as well as to a significant gain in grain yield per plant (0.9% gain per cycle) and in grain yield overall (0.8% gain per cycle) (Table 2). The Amiúdo selection cycle from the Lousada site also had a significant increase in the levels of ear moisture (0.5% gain per cycle) when compared with the initial population (Table 2). The selection performed at the Serra do Carvalho site gave rise to an Amiúdo population with a decreased percentage of stalk lodging (−1.4% gain per cycle) and to an increase in tassel branching (0.4% gain per cycle) (Table 2).

In relation to the Castro Verde population, on-farm stratified mass selection did not lead to any significant differences in the mean values of the agronomic traits evaluated in this work (Table 3). For both Amiúdo (Table 2) and Castro Verde (Table 3), no significant genotype x environment interaction was detected for the agronomic traits evaluated.

Table 2. Analysis of variance, comparison of mean values, and percentage of gain per selection cycle for the agronomic traits among Amiúdo initial population ($AM_{C0\text{-}1984}$) and selection cycles from Lousada (AM-$L_{C19\text{-}2003}$) and Serra do Carvalho (AM-$SC_{C25\text{-}2009}$).

Trait	Analysis of variance[1]			Comparison of initial population/selection cycle means[2]			% gain/cycle	% gain/cycle
	Cycle	Env	Cycle*Env	$AM_{C0\text{-}1984}$	AM-$L_{C19\text{-}2003}$	AM-$SC_{C25\text{-}2009}$	AM-$L_{C19\text{-}2003}$	AM-$SC_{C25\text{-}2009}$
Ear weight (EW), in g	***	ns	ns	124.35 b	146.07 a	162.56 a	0.9	1.2
Cob weight (CW), in g	***	*	ns	20.29 b	31.85 a	30.83 a	3.0	2.1
Cob weight/ear weight (CWEW), in g/g	***	***	ns	0.16 b	0.21 a	0.20 a	1.6	1.0
Ear moisture (EM), in %	*	***	ns	18.84 b	20.59 a	20.13 ab	0.5	–
Ear placement (E), in 1 to 9 scale[3]	ns	ns	ns	5.54 a	5.29 a	5.38 a	–	–
Leaf angle (N), in 1 to 9 scale[4]	ns	**	ns	5.42 a	5.25 a	5.29 a	–	–
Tassel branching (T), in 1 to 9 scale[5]	*	***	ns	6.21 a	6.44 ab	6.79 a	–	0.4
Root lodging (R), in %	ns	*	ns	5.48 a	7.29 a	6.32 a	–	–
Stalk lodging (S), in %	*	**	ns	9.81 a	9.53 a	6.30 b	–	-1.4
Standing plants (SP), in nº plants/ha	ns	**	ns	49,236 a	50,062 a	49,996 a	–	–
Uniformity (U), in 1 to 9 scale[6]	ns	ns	ns	3.42 a	3.58 a	3.38 a	–	–
Prolificacy (P), in nº ears/plant	ns	ns	ns	1.07 a	1.10 a	1.05 a	–	–
Grain yield (Y), in kg/ha	**	ns	ns	4,568.84 b	5,322.79 a	5,577.93 a	0.8	0.8
Grain yield per plant(YP), in g/plant	**	ns	ns	93.00 b	107.88 a	112.57 a	0.9	0.9

*[1] Significance for analysis of variance among cycles (initial population plus selection cycles), among environments (Env) and interaction between cycles and environments (Cycle*Env): ns - non-significant; * - significant at P < 0.05; ** - significant at P < 0.01; *** - significant at P < 0.001*

[2] Tukey-Kramer multiple comparisons test – mean values in each row followed by the same letter are not significantly different at P < 0.05

[3] Ear placement (E), in 1 to 9 scale: 5 indicates that the 1st ear is located in the middle of the plant; values < 5 indicates that 1st the ear is located bellow the plant middle point; values > 5 indicates that the 1st ear is located above the plant middle point

[4] Leaf angle (N), in 1 to 9 scale: 5 indicates a leaf angle = 45 °; values < 5 indicate a leaf angle <45 °; and values > 5 indicate a leaf angle > 45 °

[5] Tassel branching (T), in 1 to 9 scale: 1 indicates unbranched tassel and 9 indicates a highly branched tassel

[6] Uniformity (U), in 1 to 9 scale: 1 indicates minimum uniformity and 9 indicates maximum uniformity

Table 3. Analysis of variance and comparison of mean values for the agronomic traits among Castro Verde initial population ($CA_{C0\text{-}1994}$) and selection cycles (CA-$C_{C09\text{-}2004}$ and CA-$C_{C14\text{-}2009}$).

Trait	Analysis of variance[1]			Comparison of initial population/selection cycle means[2]		
	Cycle	Env	Cycle*Env	$CA_{C0\text{-}1994}$	CA-$C_{C09\text{-}2004}$	CA-$C_{C14\text{-}2009}$
Ear weight (EW), in g	ns	ns	ns	240.12 a	256.46 a	247.22 a
Cob weight (CW), in g	ns	**	ns	57.93 a	65.79 a	58.12 a
Cob weight/ear weight (CWEW), in g/g	ns	***	ns	0.24 a	0.26 a	0.23 a
Ear moisture (EM), in %	ns	**	ns	24.20 a	24.81 a	24.20 a
Ear placement (E), in 1-9 scale[3]	ns	*	ns	6.00 a	6.40 a	6.03 a
Leaf angle (N), in 1-9 scale[4]	ns	ns	ns	5.13 a	5.15 a	4.87 a
Tassel branching (T), in 1-9 scale[5]	ns	ns	ns	7.07 a	7.14 a	6.97 a
Root lodging (R), in %	ns	**	ns	31.99 a	31.50 a	22.53 a
Stalk lodging (S), in %	ns	***	ns	25.20 a	25.22 a	27.93 a
Standing plants (SP), in no. plants/ha	ns	**	ns	48924 a	47100 a	48403 a
Uniformity (U), in 1-9 scale[6]	ns	**	ns	3.77 a	3.80 a	3.63 a
Prolificacy (P), in no. ears/plant	ns	ns	ns	0.98 a	1.00 a	0.90 a
Grain yield (Y), in kg/ha	ns	*	ns	6862.71 a	6851.03 a	6840.93 a
Grain yield per plant (YP), in g/plant	ns	ns	ns	146.33 a	147.15 a	144.52 a

*[1] Significance for analysis of variance among cycles (initial population plus selection cycles), among environments (Env) and interaction between cycles and environments (Cycle*Env): ns - non-significant; * - significant at P < 0.05; ** - significant at P < 0.01; *** - significant at P < 0.001*

[2] Tukey-Kramer multiple comparisons test – mean values in each row followed by the same letter are not significantly different at P < 0.05

[3] Ear placement (E), in 1 to 9 scale: 5 indicates that the 1st ear is located in the middle of the plant; values < 5 indicates that 1st the ear is located bellow the plant middle point; values > 5 indicates that the 1st ear is located above the plant middle point

[4] Leaf angle (N), in 1 to 9 scale: 5 indicates a leaf angle = 45 °; values < 5 indicate a leaf angle <45 °; and values > 5 indicate a leaf angle > 45 °

[5] Tassel branching (T), in 1 to 9 scale: 1 indicates unbranched tassel and 9 indicates a highly branched tassel

[6] Uniformity (U), in 1 to 9 scale: 1 indicates minimum uniformity and 9 indicates maximum uniformity

A principal component analysis based on the agronomic data was used to summarize the multivariate changes occurring in both populations across the participatory breeding program. The first two principal components for both the Amiúdo and Castro Verde cycles retained 94.49% of the total variance, with the first component already retaining 84.37% of the observed variance (Figure 2).

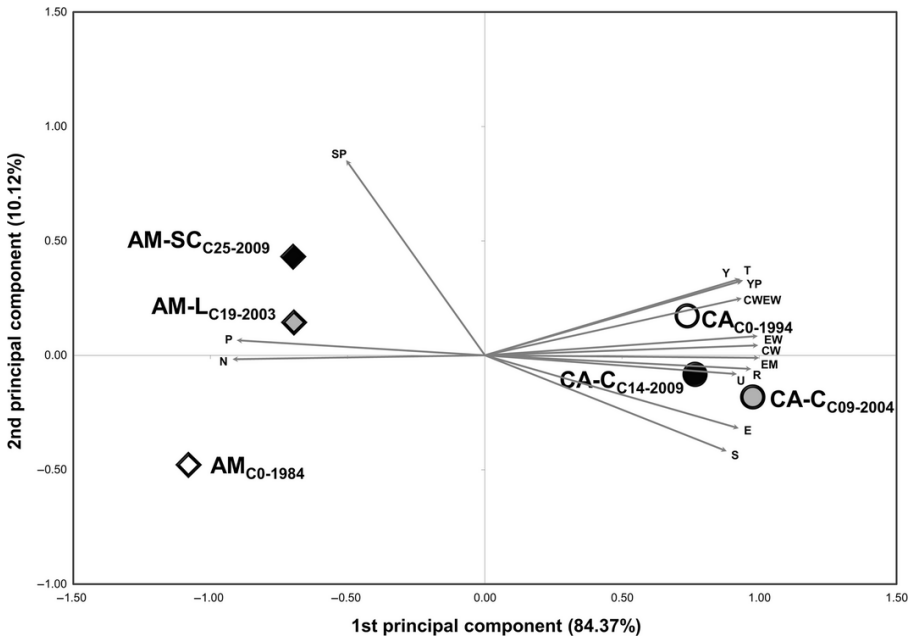

Figure 2. Biplot of principal component analysis (PCA) based on 14 agronomic traits measured in the Amiúdo cycles (initial population—$AM_{C0\text{-}1984}$; AM-$L_{C19\text{-}2003}$ selection cycle; and AM-$SC_{C25\text{-}2009}$ selection cycle) and Castro Verde cycles (initial population—$CA_{C0\text{-}1994}$; CA-$C_{C09\text{-}2004}$ selection cycle; and CA-$C_{C14\text{-}2009}$ selection cycle).

In the PCA biplot (Figure 2), the first axis clearly separated the Amiúdo from the Castro Verde populations. Moreover, for Amiúdo the first axis separated the initial population ($AM_{C0\text{-}1984}$) from the two selection cycles (AM-$L_{C19\text{-}2003}$ and AM-$SC_{C25\text{-}2009}$) in the direction of an increase in all the traits analyzed except for plant prolificacy (P) and

the angle of the leaf insertion in the stalk (N) that decreased in this direction. The second axis separated the two selection cycles, AM-$L_{C19-2003}$ and AM-$SC_{C25-2009}$, in the direction of an increase in the number of plants standing (SP), with the selection cycle from the Serra do Carvalho site having a higher number of plants standing. As for Castro Verde, and as expected by the results obtained previously for the analysis of variance (Table 3), no clear progression was observed along the selection process when comparing the position on the biplot of the initial population $CA_{C0-1994}$, the cycle from 2004 (CA-$C_{C09-2004}$), and the cycle from 2009 (CA-$C_{C14-2009}$) (Figure 2).

3.2 Molecular diversity evolution

3.2.1 Intrapopulation diversity

The molecular diversity analysis allowed tracing the overall genetic diversity evolution in the two open-pollinated populations under study. In terms of quantitative differences in the alleles detected for the Amiúdo population, 73.26% of all alleles were maintained throughout the cycles: Of the 86 alleles detected, 63 were common to all the cycles (Table S3). Only six to eight alleles (7%–9.3%), out of the 74 identified in the initial population ($AM_{C0-1984}$), were not detected in the Serra do Carvalho (AM-$SC_{C25-2009}$) and in the Lousada (AM-$L_{C19-2003}$) selection cycles, respectively (Table S2). Likewise, in terms of quantitative differences in the alleles detected for Castro Verde population, the majority of the alleles (65.91%) were maintained throughout the cycles: Of 88 alleles detected, 58 were common to all the cycles (Table S3). Only 10 alleles (11.4%), out of the 74 detected in the initial population, were not detected in the CA-$C_{C14-2009}$ selection cycle (Table S2).

As for the allelic frequencies, for both Amiúdo and Castro Verde populations a considerable proportion of the alleles detected were present in low frequencies (0.1 or less): Amiúdo cycles with 39.19% at the initial population ($AM_{C0\text{-}1984}$), 41.89% at the selection cycle from the Lousada site ($AM\text{-}L_{C19\text{-}2003}$), and 48.10% at the selection cycle from the Serra do Carvalho site (Figure S1A); and Castro Verde cycles with 47.30% at initial population ($CA_{C0\text{-}1994}$), 48.61% at the $CA\text{-}C_{C09\text{-}2004}$ selection cycle, and 50% at the $CA\text{-}C_{C14\text{-}2009}$ selection cycle (Figure S1B).

When testing for significant differences among cycles within each population in the average number of alleles detected, observed and expected heterozygosity, and inbreeding coefficients, no significant differences were observed among the cycles for both the Amiúdo and Castro Verde populations (Table 4).

The global Hardy–Weinberg equilibrium test detected a significant departure from Hardy–Weinberg equilibrium in the Amiúdo cycle, $AM\text{-}SC_{C25\text{-}2009}$, and in the Castro Verde cycle, $CA\text{-}C_{C14\text{-}2009}$, both due to heterozygote deficiency (F_{IS} = 0.042, p-value <.01; and F_{IS} = 0.082, p-value <.05, respectively) (Table 4). When testing for the departure from Hardy–Weinberg equilibrium by individual locus in both the Amiúdo and Castro Verde populations, the majority of the loci had their genotypic frequencies in accordance with Hardy–Weinberg expectations (Table S4).

With the objective of testing for temporal changes in the allele frequencies distribution, the effective population size (N_e) was estimated by a temporal-based method under sample plan II. For Amiúdo, the estimated effective population size for the Lousada site was N_e = 119.6, while for the Serra do Carvalho site the N_e value was bigger (N_e = 243.7) (Table S5). For Castro Verde, the estimated effective population size was N_e = 161.7 (Table S5). After a

Bonferroni multiple-test correction, no significant temporal variation of allele frequencies was detected for both populations and selection sites (Amiúdo: Table S6; Castro Verde: Table S7).

Table 4. Genetic variability estimates for Amiúdo initial population ($AM_{C0\text{-}1984}$) and Castro Verde initial population ($CA_{C0\text{-}1994}$) and derived selection cycles.

Population / Selection cycle	N	N_{av}	N_{pr}	H_O	H_E	F_{IS}	P-value HWE
$AM_{C0\text{-}1984}$	30	3.70	3	0.537	0.532	-0.009	ns
AM-$L_{C19\text{-}2003}$	30	3.70	1	0.523	0.531	0.014	ns
AM-$SC_{C25\text{-}2009}$	30	3.95	4	0.503	0.526	0.042	**
P-value*(KW)		0.961		0.584	0.725	0.520	
$CA_{C0\text{-}1994}$	30	3.70	4	0.482	0.482	0.000	ns
CA-$C_{C09\text{-}2004}$	30	3.60	2	0.456	0.482	0.054	ns
CA-$C_{C14\text{-}2009}$	30	3.80	6	0.457	0.498	0.082	*
P-value*(KW)		0.911		0.790	0.930	0.825	

** P-value of Kruskal-Wallis test among cycles (initial populations and derived selection cycles).*

*N: number of individuals, N_{av}: average number of alleles, N_{pr}: number of private alleles, H_O: observed heterozygosity, H_E: gene diversity or expected heterozygosity, F_{IS}: inbreeding coefficient, P-value HWE: The probability global test for Hardy-Weinberg equilibrium (HWE) for each cycle was based on Markov chain method. ns - non-significant; * - significant at P < 0.05; ** - significant at P < 0.01*

3.2.2 Differentiation among cycles

The genetic differentiation among cycles within each population was tested following the framework of AMOVA. The AMOVA results showed that for the Amiúdo population, the percentage of variance that could be attributed to differences among all cycles represented 2.86% of the total molecular variation (Table 5). The pairwise comparisons between Amiúdo cycles showed that stratified mass selection led overall to a significant but small genetic differentiation (given the significant ϕ_{ST} values; Table 5). For the Castro Verde population, AMOVA showed that the variation among all cycles represented only 1.72% of the total molecular variation (Table 5). In this case, stratified mass selection did not generate a significant

genetic differentiation between $CA_{C0\text{-}1994}$ and $CA\text{-}C_{C09\text{-}2004}$ (ϕ_{ST} = 0.003, *p*-value >.05) (Table 5).

Table 5. Analysis of molecular variance (AMOVA) results for the partitioning of SSR variation among and within Amiúdo cycles ($AM_{C0\text{-}1984}$, $AM\text{-}L_{C19\text{-}2003}$, and $AM\text{-}SC_{C25\text{-}2009}$) and Castro Verde cycles ($CA_{C0\text{-}1994}$, $CA\text{-}C_{C09\text{-}2004}$, and $CA\text{-}C_{C14\text{-}2009}$).

Comparison	% Total variance		ϕ-statistics[1]	P(ϕ)[2]
	Among Cycles	Within Cycles		
$AM_{C0\text{-}1984}$ *vs.* $AM\text{-}L_{C19\text{-}2003}$	4.33	95.67	0.043	***
$AM_{C0\text{-}1984}$ *vs.* $AM\text{-}SC_{C25\text{-}2009}$	2.98	97.02	0.030	***
$AM\text{-}L_{C19\text{-}2003}$ *vs.* $AM\text{-}SC_{C25\text{-}2009}$	1.22	98.78	0.012	*
All Amiúdo cycles	2.86	97.14	0.029	***
$CA_{C0\text{-}1994}$ *vs.* $CA\text{-}C_{C09\text{-}2004}$	0.34	99.66	0.003	ns
$CA_{C0\text{-}1994}$ *vs.* $CA\text{-}C_{C14\text{-}2009}$	2.40	97.60	0.024	***
$CA\text{-}C_{C09\text{-}2004}$ *vs.* $CA\text{-}C_{C14\text{-}2009}$	2.36	97.64	0.024	***
All Castro Verde cycles	1.72	98.28	0.017	***

[1] ϕ-statistics: corresponds to an analogous to the fixation index (F_{ST}) which measures the degree of genetic differentiation among populations/selection cycles (ϕ_{ST})

*[2] P(ϕ): the level of significance of the ϕ-statistics was tested by non-parametric randomization tests using 10,000 permutations. ns - non-significant; * - significant at P < 0.05; *** - significant at P < 0.001*

3.2.3 Genetic relationships among individuals

The factorial correspondence analysis depicts graphically the genetic proximity/differentiation within and among initial populations and selection cycles. From the factorial correspondence analysis of the Amiúdo population, the first axis, which accounted for 66.16% of the observed genotypic variance, separated the initial population ($AM_{C0\text{-}1984}$) from its selection cycles. The second axis, which accounted for 33.84% of the observed genotypic variance, separated the selection cycle from the Lousada site ($AM\text{-}L_{C19\text{-}2003}$) from the selection cycle from the Serra do Carvalho site ($AM\text{-}SC_{C25\text{-}2009}$; Figure 3). From the factorial correspondence analysis of Castro Verde, the first axis, which accounted for 63.85% of the observed

genotypic variance, separated the most recent selection cycle (CA-$C_{C14-2009)}$ from the other two. The second axis, which accounted for 36.15% of the observed genotypic variance, separated the initial cycle ($CA_{C0-1994}$) from the 2004 selection cycle (CA-$C_{C09-2004}$; Figure 4).

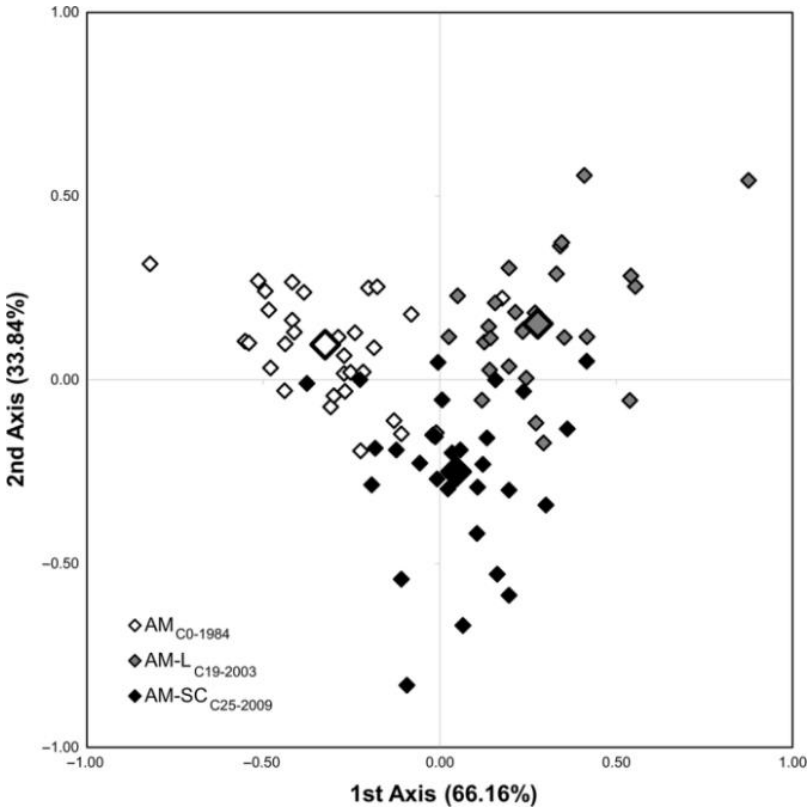

Figure 3. Factorial correspondence analysis (FCA) of 90 maize plants belonging to the Amiúdo cycles (initial population − $AM_{C0-1984}$; AM-$L_{C19-2003}$ selection cycle; and AM-$SC_{C25-2009}$ selection cycle). Each individual genotype is indicated by a small symbol, while the cycle's mean value is represented by larger ones.

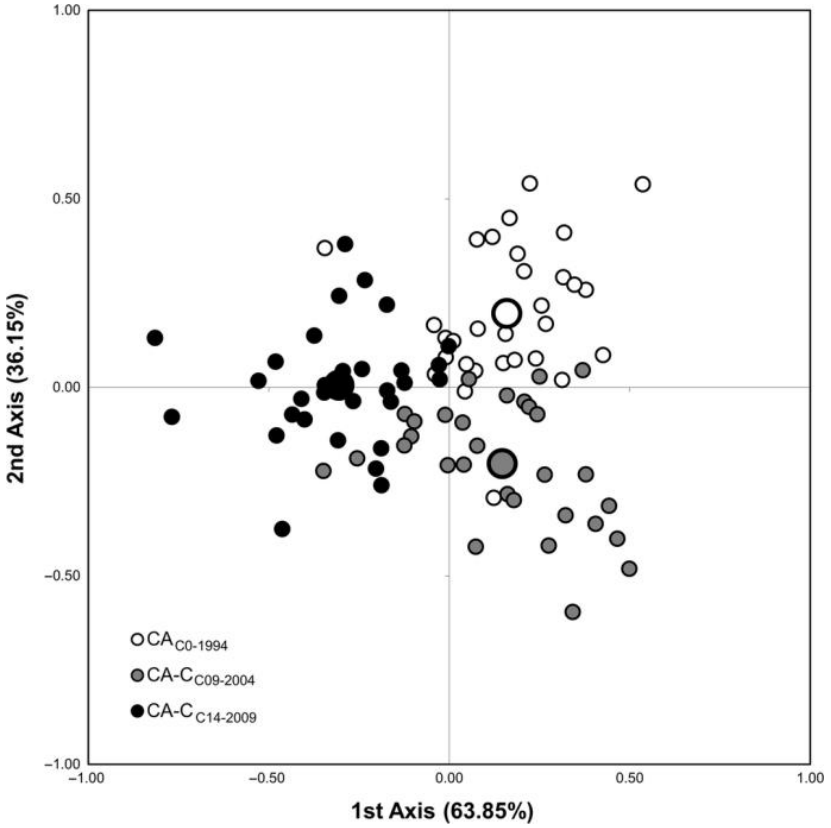

Figure 4. Factorial correspondence analysis (FCA) of 90 maize plants belonging to the Castro Verde cycles (initial population − $CA_{C0\text{-}1994}$; CA-$C_{C09\text{-}2004}$ selection cycle; and CA-$C_{C14\text{-}2009}$ selection cycle). Each individual genotype is indicated by a small symbol, while the cycle's mean value is represented by larger ones.

3.3 Quality evolution

In relation to Amiúdo quality evaluation, the breeding activities led, in the material developed both at Lousada (AM-$L_{C19\text{-}2003}$ cycle) and at Serra do Carvalho (AM-$SC_{C25\text{-}2009}$ cycle), to a slight increase in the total carotenoid content (TCC) and in the color red/green index (*a**), accompanied by a decrease in the levels of γ-tocopherol (GT), protein (PR), fiber (FI), total volatile aldehydes (AL), total free

phenolic (PH) compounds, *p*-coumaric acid (CU), and ferulic acid (FE) (Table S8).

In the case of Castro Verde quality evaluation, although the results showed first a reduction of the flour's yellowness (taken as color parameter *b** values) from $CA_{C0\text{-}1994}$ to CA-$C_{C09\text{-}2004}$ and afterward from CA-$C_{C09\text{-}2004}$ to CA-$C_{C14\text{-}2009}$ cycle, the *b** value stopped decreasing. Moreover, it was observed an increase in the levels of (α-, δ-, and γ-) tocopherols (AT, DT, GT), and *p*-coumaric acid (CU), as well as a decrease in the levels of fiber (FI), protein (PR), and total free phenolic (PH) compounds along the selection cycles. Nevertheless, for Castro Verde, the majority of the quality traits (10 of 18) variation was erratic along selection cycles.

As for the principal component analysis based on the quality data in both the Amiúdo and Castro Verde populations, the first two components retained 73.20% of the total observed variance, with the first component explaining 50.99% of the observed variance (Figure 5). The traits that primarily influenced the first component were α- and δ-tocopherol (AT and DT), fat (FT), peak and trough viscosities (PV and TV), and protein content (PR). The trait that primarily influenced the second component was the *p*-coumaric acid (CU) content.

The PCA biplot revealed an increase in the levels of α- and δ-tocopherol (AT and DT) and fat (FT) when comparing the Amiúdo initial population ($AM_{C0\text{-}1984}$) with the Amiúdo cycle from the Lousada selection site (AM-$L_{C19\text{-}2003}$). While comparing the Amiúdo initial population ($AM_{C0\text{-}1984}$) with the Amiúdo cycle from the Serra do Carvalho selection site (AM-$L_{C25\text{-}2009}$), an opposite trend was depicted with a decrease in the levels of α- and δ-tocopherol (AT and DT), and fat (FT), accompanied by a decrease in levels of *p*-coumaric acid (CU).

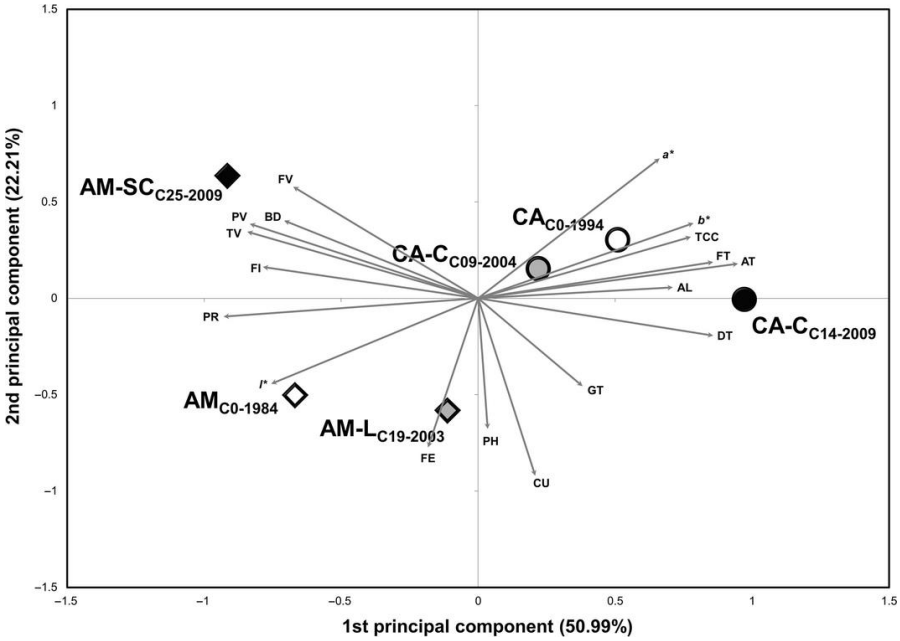

Figure 5. Biplot of principal component analysis (PCA) based on 18 quality traits in the Amiúdo cycles (initial population—$AM_{C0-1984}$; AM-$L_{C19-2003}$ selection cycle; and AM-$SC_{C25-2009}$ selection cycle) and Castro Verde cycles (initial population—$CA_{C0-1994}$; CA-$C_{C09-2004}$ selection cycle; and CA-$C_{C14-2009}$ selection cycle).

4 Discussion

Amiúdo and Castro Verde are two historical open-pollinated maize populations that have been subjected to on-farm stratified mass selection, in the context of a long-term participatory breeding program. The results presented here revealed that this participatory program is improving or maintaining yield and quality parameters while preserving the genetic diversity of maize populations. Additionally, this program is empowering farmers as they keep the decision power and learn some basic population improvement methodologies, and at the same time represents an alternative strategy for endangered genetic resources' on-farm conservation.

4.1 Phenotypic effects of stratified mass selection

The results obtained from multi-location field trials, established to evaluate the effects of stratified mass selection in these two maize populations, showed that this methodology was able to improve the Amiúdo population, according to the established selection criteria in two different selection sites (Lousada and Serra do Carvalho). Nevertheless, according to the data collected, the same methodology failed to lead to an agronomic improvement of the Castro Verde population.

The Amiúdo population, integrated on the PPB program since its beginning, was selected by two different people, in two different selection sites, but with similar edaphic–climatic conditions. For both selection sites, achieving a higher-yielding population was the breeding objective established by the farmer. Indeed, Amiúdo population had a yield increase through mass selection (0.8% gain per cycle) accompanied by heavier cobs and ears. This gain was, however, inferior to the experimental values obtained across long-term maize recurrent selection methods for population improvement, as reviewed by Betrán et al. (2004). According to Betrán et al. (2004), when grain yield is the primary selection criterion, mass selection showed on average a 1.8% gain per cycle, being this value often smaller than the average values obtained with family-based recurrent selection, such as selfed—S1 or S2—family selection (with 7% and 5% gain per cycle, respectively). One of the reasons for the slower yield progress observed in Amiúdo population in comparison with these reviewed values, besides its particular genetic background, may be a reflection of the lower selection intensity applied under the present participatory program (1%–5%).

As for Castro Verde population, the phenotypic data showed that stratified mass selection was able to partially induce phenotypic differences that follow the direction of the breeding objectives (maintenance of orange grain color set as breeding criterion after 2001). Nevertheless, an analysis of most of the other breeding criteria—achieve bigger ears, decrease the height of the ear insertion in the plants, and increase stalk resistance—showed that no significant improvements were obtained for the Castro Verde population using this methodology.

4.2 Implications for a quality-oriented breeding program

An important aspect of both the Amiúdo and Castro Verde populations is the fact that their flours can be used for food. In fact, a recent sensory hedonic analysis of maize bread, including bread obtained from these populations, showed that both populations were able to produce bread with preferential characteristics (Carbas et al., 2016). With the objective of integrating these two populations in a quality-oriented breeding program in due course, several traits related to consumer preferences and technological, nutritional, and organoleptic properties (quality traits) were measured. It was observed that the majority of those traits progressed erratically along the breeding program for the Castro Verde population. One exception was the total carotenoid content, which can be selected efficiently by choosing the more yellow/orange ears as the *b** parameter (yellowness) is highly correlated with total carotenoid content (Kljaka et al., 2014). In general for quality traits, as the ones considered in this work, a direct visual selection, like the one performed for the agronomic traits, is not possible, and other complementary breeding methodologies are needed to encourage their effective improvement by farmers.

4.3 Breeding program weaknesses and strengths analysis

When grain yield was the primary breeding objective, on-farm stratified mass selection, as described in this work, was effective in improving population yield although at a slower rate than what can be obtained through other more complex family-based recurrent selection methods. With more diverse breeding objectives, as in the case of Castro Verde population, the stratified mass selection was not always effective in achieving the same progress.

An extensive compilation of several cases of yield improvement achieved through mass selection in maize can be found at Hallauer et al. (2010, table 7.8, therein). A few examples that show the potential of stratified mass selection specifically in the context of a participatory maize breeding program were described in Mendes-Moreira et al. (2008, 2009) and Smith et al. (2001). In the first two works, two other maize populations from the same Portuguese breeding program as in the present study had their agronomic performance improved in line with the farmers' breeding objectives (Mendes-Moreira et al., 2008, 2009). Also Smith et al. (2001) showed that tree cycles of stratified mass selection applied to five different Mexican maize populations were sufficient to obtain an increase in yield. Several factors have been identified as having an impact on mass selection effectiveness or ineffectiveness (Hallauer et al., 2010). Among them, one can highlight the trait under selection, an adequate isolation, the sample size utilized, genotype x environment interaction, and the precision of the experimental techniques used (environmental control, parental control). In the present work, it was shown that the selection methodology was able to alter traits related to ear architecture in the Amiúdo population, and therefore, the lack of agronomic progress in ear architecture-related traits in the Castro Verde population should

not be due to the trait under selection *per se*. Moreover, as the analysis of variance did not detect a significant genotype-by-environment interaction, the lack of Castro Verde progress should not be a consequence of this interaction. Instead, it could be most likely related to two particular aspects of the Castro Verde population: First, as the selection criterion until the year 2000 was set to get bigger ears, one hypothesis is that because this population had already ears of a significant size before entering the breeding program, the farmer was not fully engaged with the breeding activities. Second, after 2001, this population started to be selected at Coimbra site by the breeder. Therefore, another hypothesis for the lack of observable agronomic progress is that the population did not have adequate isolation, as other populations were also being grown at the same site; and the number of individual plants screened may have been too small to select/capture the best genotypes. Indeed, Castro Verde initial population, which resulted from years of farmers traditional selection based mainly on ear traits evaluated after harvest, had already a high grain yield for an open-pollinated maize population (6,862.71 kg/ha). Probably due to this, a yield increase was not the main objective of the farmer involved on Castro Verde selection. This, however, was not the case for the farmer involved on Amiúdo selection that was aiming to improve the population initial yield (4,568.84 kg/ha). Nevertheless, both original maize populations showed on average higher yields than the only data publicly available on nonimproved historical Portuguese maize populations with high quality potential for maize bread *broa* production (Vaz Patto et al., 2007). Grain yield of these traditional populations was evaluated in a common-garden field experiment, and it varied from 755 to 3,757 kg/ha, with an average of 1,982 kg/ha (Vaz Patto et al., 2007).

In the maize populations analyzed in the present study, not only natural selection but also human selection is affecting yield. In a review by Murphy et al. (2013), several examples of the effectiveness of evolutionary breeding (accounting only for natural selection) in improving the agronomic fitness of self-pollinated cereal crops have been examined. With this breeding approach, improvement resulted from natural selection favoring high-yielding genotypes as an outcome of the relationship between the yield capacity of an individual plant and its fitness components (Murphy et al., 2013). This yield increase is highly dependent on the selective environmental pressure and may affect maturity, plant height, and relationships among agronomic important traits unfavorably (Phillips & Wolfe, 2005). A comparison between the yield progress attained under the studied participatory breeding program and the yield progress that might be attained with an evolutionary breeding approach could have generated relevant information on the effectiveness of the human (artificial) selection versus natural selection. Unfortunately, no references were found in the literature on the effect of evolutionary breeding in maize populations to allow a direct comparison with the present study. However, by performing the selection of Amiúdo and Castro Verde populations within the target environment (at the farmers' fields), on-farm participatory breeding guarantees local adaptation and it may also counteract undesirable changes caused by natural selection in traits of agronomic importance. Moreover, by respecting farmers' breeding objectives, an increase in the ratio of improved populations adopted by the farmer can be obtained.

Although one can argue that differences in response to selection in a similar genetic background may be due to different intensity or accuracy of selection, the acceptance and the enthusiasm of the farmers to join the program are the best guarantees of success.

Farmers need to be fully engaged in the selection decision process (breeding objectives) but be open to accept breeder recommendations (preharvest parental control + postharvest selection).

One open question in the present study is: How able is the farmer to perform pre-harvest trait selection? In the present work, the preharvest selection was not exclusive but mainly performed by the breeder, and therefore, the farmer's ability could not be clearly evaluated. Nevertheless, theoretically, the preharvest selection methodologies proposed in the Portuguese participatory breeding program are very straightforward and are beforehand demonstrated by the breeder in the farmer's field. Therefore, these methodologies should be easily implemented by any farmer engaged in the breeding process. Indeed, it has been already demonstrated by Mendes-Moreira et al. (2008) that such preharvest methodologies were successfully implemented by farmers in another maize population from the same participatory breeding program. The farmer's motivation and time availability/field dimensions (the bigger the field, the larger amount of time needed for stratified preharvest selection) seem to be the two main limitations for the successful implementation of this preharvest selection.

4.4 Genotypic effects of stratified mass selection

The effect of stratified mass selection in the genetic diversity levels of the two populations was also evaluated using SSRs. This analysis showed that the overall genetic diversity was maintained in both populations. In particular, even in the Amiúdo population where phenotypic modifications on ear morphology and yield gain were detected, no significant changes were identified on the overall genetic diversity levels, measured by the average number of alleles detected,

observed and expected heterozygosity, and inbreeding coefficients. Also, no significant temporal variation of allele frequencies was detected in any of populations under study, indicating that the observed differences in allele frequency are more likely a result of genetic drift and/or sampling error (Waples, 1989a). As opposed to the results obtained by Labate et al. (1999) and Solomon et al. (2010), in which the authors detected a loss of genetic diversity in maize population subjected to few as 11 and 12 cycles of reciprocal recurrent selection, no significant differences in genetic diversity levels were identified in the current study. According to Hoban et al. (2014), changes in genetic diversity levels are most likely identified only when the effective population size is smaller than 100 individuals. In the present work, both populations had an effective population size bigger than 100, by contrast to the smaller effective population sizes estimated for the maize populations in Labate et al. (1999) and Solomon et al. (2010). In addition, the results presented here concur with the results previously described for the Portuguese Pigarro maize population (Vaz Patto et al., 2008) where stratified mass selection demonstrated to be an effective way to conserve diversity on-farm, and at the same time allowed relevant phenotypic improvements to be achieved.

4.5 Final remarks

In conclusion, on-farm stratified mass selection in the context of a participatory plant breeding program was shown to improve the agronomic performance of the Amiúdo population selected in two different selection sites. Moreover, for both the Amiúdo and Castro Verde populations, the breeding activities retained the populations' genetic diversity. The unpredictability of the evolution of quality parameters along this breeding program also brings to light the need

to develop efficient selection tools to maintain or improve these traits. Molecular markers associated with those traits and/or high throughput spectroscopy-based phenotypic screening methodologies are among the tools that may aid in the improvement of characteristics that cannot be easily (visually) selected by farmers. The implementation of such breeding tools into participatory selection brings up another issue: To make these tools easily available, a platform of participatory research connecting enthusiastic, open-minded farmers, breeders, and scientists must be built to make its application a reality.

Acknowledgments

MLA performed the DNA isolation and the SSR genotyping, the analysis of the molecular and phenotypic data, and drafted the manuscript. CB and MP participated in the establishment of the field trials and performed the agronomic characterization. BC and MB performed the quality characterization. PMM coordinated the field trials, performed the agronomic characterization, and participated in the manuscript revision. CB and MRB coordinated and participated in the acquisition of quality data, and participated in the manuscript revision. ZS participated in the analysis of the molecular and phenotypic data and participated in the manuscript revision. MCVP designed and coordinated the study and participated in the drafting and revising of the manuscript.

The research leading to these results has received funding from the European Union's 7th Framework Programme (FP7/2007-2013) and Horizon 2020 Framework Programme under grant agreement No. 245058 and No. 633571, respectively, and from the Fundação para a Ciência e a Tecnologia through R&D unit, UID/Multi/04551/2013 (GreenIT). MLA and MCVP Vaz Patto were supported by Fundação para a Ciência e a Tecnologia (SFRH/BD/77723/2011; FCT

Investigator 2015 contract IF/01337/2014, respectively). The authors would like to acknowledge Doctor Silas Pego and the farmers that were involved in the development of the maize populations, as well as the Portuguese Bank of Plant Germplasm for providing the seeds from the initial landraces and selection cycles used on this work. The authors would also like to acknowledge all the anonymous reviewers that contributed to the discussion and improvement of this work.

References

AACC International (2012). Approved methods of analysis Method 14-60.01, 11th ed. Total carotenoid content of cereal grains and flours. St. Paul, MN, USA: AACC International.

Belkhir, K., Borsa, P., Chikhi, L., Raufaste, N., & Bonhomme, F. (2004). GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).

Betrán, J., Bänziger, M., & Menz, M. (2004). Corn breeding. In C. Wayne Smith, J. Betrán, & E. C. A. Runge (Eds.), *Corn: Origin, history, technology, and production* (pp. 305–398). New York, NY: Wiley.

Brites, C., Trigo, M. J., Santos, C., Collar, C., & Rosell, C. M. (2010). Maize-based gluten-free bread: Influence of processing parameters on sensory and instrumental quality. *Food Bioprocess Technology*, 3, 707–715.

Carbas, B., Vaz Patto, M. C., Bronze, M. R., Bento da Silva, A., Trigo, M. J., & Brites, C. (2016). Maize flour parameters that are related to the consumer perceived quality of 'broa' specialty bread. *Food Science and Technology*, 36, 259–267.

Ceccarelli, S. (2015). Efficiency of plant breeding. *Crop Science*, 55, 87–97.

Ceccarelli, S., Al-Yassin, A., Goldringer, I., Mendes-Moreira, P., & Chable, V. (2012). *Analysis of major PPB experiences worldwide*. Retrieved from http://www.solibam.eu/modules/wfdownloads/singlefile.php?cid=12&lid=33

Ceccarelli, S., Galie, A., & Grando, S. (2013). Participatory breeding for climate change-related traits. In C. Kole (Ed.), *Genomics and breeding for climate*

resilient crops. Concepts and strategies. Volume 1 (pp. 331–376). Heidelberg, Germany: Springer.

Ceccarelli, S., Grando, S., Maatougui, M., Michae, L. M., Slash, M., Haghparast, R., et al. (2010). Plant breeding and climate changes. *Journal of Agricultural Science*, 148, 627–637.

Cleveland, D. A., Soleri, D., & Smith, S. E. (1999). Farmer plant breeding from biological perspective: Implications for collaborative plant breeding. CIMMYT Economics Working Paper CIMMYT, Mexico DF: CIMMYT, 99–10.

Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, 14, 209–214.

Excoffier, L., Laval, G., & Schneider, S. (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics* Online, 1, 47–50.

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction sites. *Genetics*, 131, 479–491.

Fu, Y. B. (2015). Understanding crop genetic diversity under modern plant breeding. *Theoretical and Applied Genetics*, 128, 2131–2142.

Gardner, C. O. (1961). An evaluation of the effects of mass selection and seed irradiation with thermal neutrons on yield of corn. *Crop Science*, 1, 241–245.

Guo, S. W., & Thompson, E. A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, 48, 361–372.

Hallauer, A. R. (1999). Conversion of tropical germplasm for temperate area use. Illinois Corn Breeders School, 35, 20–36.

Hallauer, A. R., Carena, M. J., & Miranda Filho, J. D. (2010). Quantitative genetics in maize breeding (Vol. 6). New York, NY: Springer.

Hellin, J., Bellon, M. R., & Hearne, S. J. (2014). Maize landraces and adaptation to climate change in Mexico. *Journal of Crop Improvement*, 28, 484–501.

Hoban, S., Arntzen, J. A., Bruford, M. W., Godoy, J. A., Rus Hoelzel, A., Segelbacher, G., et al. (2014). Comparative evaluation of potential indicators and temporal sampling protocols for monitoring genetic erosion. *Evolutionary Applications*, 7, 984–998.

ISO 9936 (2006). Animal and vegetable fats and oils – Determination of tocopherol and tocotrienol contents by high-performance liquid chromatography.

Jarvis, D. I., Hodgkin, T., Sthapit, B. R., Fadda, C., & Lopez-Noriega, I. (2011). An heuristic framework for identifying multiple ways of supporting the conservation and use of traditional crop varieties within the agricultural production system. *Critical Reviews in Plant Sciences*, 30, 125–176.

Khoury, C. K., Bjorkman, A. D., Dempewolf, H., Ramirez-Villegas, J., Guarino, L., Jarvis, A., et al. (2014). Increasing homogeneity in global food supplies and the implications for food security. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 4001–4006.

Klensporf, D., & Jelén, H. H. (2005). Analysis of volatile aldehydes in oat flakes by SPME-GC/MS. *Polish Journal of Food and Nutrition Sciences*, 14, 389–395.

Kljaka, K., Grbešaa, D., & Karolyib, D. (2014). Reflectance colorimetry as a simple method for estimating carotenoid content in maize grain. *Journal of Cereal Science*, 59, 109–111.

Labate, J. A., Lamkey, K. R., Lee, M., & Woodman, W. L. (1999). Temporal changes in allele frequencies in two reciprocally selected maize populations. *Theoretical and Applied Genetics*, 99, 1166–1178.

Lançon, J., Pichaut, J. P., Djaboutou, M., Lewicki-Dhainaut, S., Viot, C., & Lacape, J. M. (2008). Use of molecular markers in participatory plant breeding: Assessing the genetic variability in cotton populations bred by farmers. *Annals of Applied Biology*, 152, 113–119.

Lawrence, C. J., Harper, L. C., Schaeffer, M. L., Sen, T. Z., Seigfried, T. E., & Campbell, D. A. (2008). MaizeGDB: The maize model organism database for basic, translational, and applied research. *International Journal of Plant Genomics*, 2008, 496957.

Lipka, A. E., Gore, M. A., Magallanes-Lundback, M., Mesberg, A., Lin, H., Tiede, T., et al. (2013). Genome-wide association study and pathway-level

analysis of tocochromanol levels in maize grain. *G3: Genes|Genomes|Genetics*, 3, 1287–1299.

Liu, R. H. (2003). Health benefits of fruit and vegetables are from additive and synergistic combinations of phytochemicals. *American Journal of Clinical Nutrition*, 78, 517S–520S.

Lopez-Martinez, L. X., Oliart-Ros, R. M., Valerio-Alfaro, G., Lee, C. H., Parkin, K. L., & Garcia, H. S. (2009). Antioxidant activity, phenolic compounds and anthocyanins content of eighteen strains of Mexican maize. *Food Science and Technology*, 42, 1187–1192.

Mendes-Moreira, P. M. R., Pego, S. E., Vaz Patto, C., & Hallauer, A. R. (2008). Comparison of selection methods on 'Pigarro', a Portuguese improved maize population with fasciation expression. *Euphytica*, 163, 481–499.

Mendes-Moreira, P. M. R., Vaz Patto, M. C., Mota, M., Mendes-Moreira, J., Santos, J. P. N., Santos, J. P. P., et al. (2009). 'Fandango': Long term adaptation of exotic germplasm to a Portuguese on-farm-conservation and breeding project. *Maydica*, 54, 269–285.

Murphy, K. M., Carter, A. H., & Jones, S. S. (2013). Evolutionary breeding and climate change. In C. Kole (Ed.), Genomics and breeding for climate resilient crops. Concepts and strategies. Volume 1 (pp. 377–389). Heidelberg, Germany: Springer.

Nei, M., & Tajima, F. (1981). Genetic drift and estimation of effective population size. *Genetics*, 98, 625–640.

Ortiz, R. (2011). Agrobiodiversity management for climate change. In J. M. Lenné, & D. Wood (Eds.), *Agrobiodiversity management for food security* (pp. 189–211). Wallingford, Oxon, United Kingdom: CAB International.

Owens, B. F., Lipka, A. E., Magallanes-Lundback, M., Tiede, T., Diepenbrock, C. H., Kandianis, C. B., et al. (2014). A foundation for provitamin A biofortification of maize: Genome-wide association and genomic prediction models of carotenoid levels. *Genetics*, 198, 1699–1716.

Phillips, S. L., & Wolfe, M. S. (2005). Evolutionary plant breeding for low input systems. *The Journal of Agricultural Science*, 143, 245–254.

Pingali, P. L. (ed.). (2001). CIMMYT 1999–2000 World Maize Facts and Trends. Meeting World Maize Needs: Technological Opportunities and Priorities for the Public Sector Mexico. D.F.: CIMMYT.

Rauf, S., Teixeira da Silva, J. A., Khan, A. A., & Naveed, A. (2010). Consequences of plant breeding on genetic diversity. International *Journal of Plant Breeding*, 4, 1–21.

Raymond, M., & Rousset, F. (1995). GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, 86, 248–249.

Rether, B., Delmas, G., & Laouedj, A. (1993). Isolation of polysaccharide-free DNA from plants. Plant Molecular Biology Reporter, 11, 333–337.

Rice, W. R. (1989). Analyzing tables of statistical tests. *Evolution*, 43, 223–225.

Saghai-Maroof, M. A., Soliman, K. M., Jorgensen, R. A., & Allard, R. W. (1984). Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 81, 8014–8018.

SAS Institute Inc (2004). The SAS system for windows, V8. 02. Cary, NC: SAS Institute.

Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, 18, 233–234.

Silva, S., Bronze, M. R., Figueira, M. E., Siwy, J., Mischak, H., Combet, E., et al. (2015). Impact of a 6-wk olive oil supplementation in healthy adults on urinary proteomic biomarkers of coronary artery disease, chronic kidney disease, and diabetes (types 1 and 2): A randomized, parallel, controlled, double-blind study. *The American Journal of Clinical Nutrition*, 101, 44–54.

Silva, S., Gomes, L., Leitão, F., Coelho, A. V., & Vilas Boas, L. (2006). Phenolic compounds and antioxidant activity of Olea europaea L. fruits and leaves. *Food Science Technology International*, 12, 385–396.

Singleton, V. L., Orthofer, R., & Lamuela-Raventos, R. M. (1999). Analysis of total phenols and other oxidant substrates and antioxidants by means of Folin-Ciocalteu reagent. *Methods in Enzymology*, 299, 157–178.

Smith, M. E., Castillo, F. G., & Gómez, F. (2001). Participatory plant breeding with maize in Mexico and Honduras. *Euphytica*, 122, 551–565.

Solomon, K. F., Martin, I., & Zeppa, A. (2010). Temporal genetic structure patterns in tropical maize populations under reciprocal recurrent selection. *Euphytica*, 176, 239–249.

Vaz Patto, M. C., Mendes-Moreira, P. M., Alves, M. L., Mecha, E., Brites, C., Bronze, R., et al. (2013). Participatory plant quality breeding: An ancient art revisited by knowledge sharing. The Portuguese experience. In S. B. Andersen (Ed.), *Plant breeding from laboratories to fields* (pp. 255–288). Rijeka: InTech.

Vaz Patto, M. C., Moreira, P. M., Almeida, N., Satovic, Z., & Pego, S. (2008). Genetic diversity evolution through participatory maize breeding in Portugal. *Euphytica*, 161, 283–291.

Vaz Patto, M. C., Moreira, P. M., Carvalho, V., & Pego, S. (2007). Collecting maize (Zea mays L. convar. mays) with potential technological ability for bread making in Portugal. *Genetic Resources and Crop Evolution*, 54, 1555–1563.

Waples, R. S. (1989a). Temporal variation in allele frequencies: Test the right hypothesis. *Evolution*, 43, 1236–1251.

Waples, R. S. (1989b). A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics*, 95, 489–502.

Wheeler, T., & von Braun, J. (2013). Climate change impacts on global food security. *Science*, 341, 508–513.

Supplementary material

Supplementary information is available online through the link http://onlinelibrary.wiley.com/store/10.1111/eva.12549/asset/supinfo/eva12549-sup-0001-Supinfo.docx?v=1&s=2a0fa56f032cd7dc9c885c657d5123b54e20740d, with the file name: eva12549-sup-0001-Supinfo.docx

Chapter III

Setting up decision-making tools for a quality-oriented participatory maize breeding program

The work presented in this chapter was published in the following research publication:

Alves M. L., Brites C., Paulo M., Carbas B., Belo M., Mendes-Moreira P., Brites C., Bronze M. R., Gunjača J., Šatović Z., & Vaz Patto M. C. (2017) Setting up decision-making tools towards a quality-oriented participatory maize breeding program. *Frontiers in Plant Science*, 8, 2203. doi: 10.3389/fpls.2017.02203

In this research paper, Mara Lisa Alves performed the DNA isolation and the SSR genotyping, the analysis of molecular and quality data, and drafted the manuscript. (See Acknowledgements section for authors' contributions)

Abstract

Previous studies have reported promising differences in the quality of kernels from farmers' maize populations collected in a Portuguese region known to produce maize-based bread. However, several limitations have been identified in the previous characterizations of those populations, such as a limited set of quality traits accessed and a missing accurate agronomic performance evaluation. The objectives of this study were to perform a more detailed quality characterization of Portuguese farmers' maize populations; to estimate their agronomic performance in a broader range of environments; and to integrate quality, agronomic, and molecular data in the setting up of decision-making tools for the establishment of a quality-oriented participatory maize breeding program.

Sixteen farmers' maize populations, together with 10 other maize populations chosen for comparison purposes, were multiplied in a common-garden experiment for quality evaluation. Flour obtained from each population was used to study kernel composition (protein, fat, fiber), flour's pasting behavior, and bioactive compound levels (carotenoids, tocopherols, phenolic compounds). These maize populations were evaluated for grain yield and ear weight in nine locations across Portugal; the populations' adaptability and stability were evaluated using additive main effects and multiplication interaction (AMMI) model analysis. The phenotypic characterization of each population was complemented with a molecular characterization, in which 30 individuals per population were genotyped with 20 microsatellites.

Almost all farmers' populations were clustered into the same quality-group characterized by high levels of protein and fiber, low levels of carotenoids, volatile aldehydes, α- and δ-tocopherols, and

breakdown viscosity. Within this quality-group, variability in particular quality traits (color and some bioactive compounds) could still be found. Regarding the agronomic performance, farmers' maize populations had low but considerably stable grain yields across the tested environments. As for their genetic diversity, each farmers' population was genetically heterogeneous; nonetheless, all farmers' populations were distinct from each other's.

In conclusion, and taking into consideration different quality improvement objectives, the integration of the data generated within this study allowed the outline and exploration of alternative directions for future breeding activities. As a consequence, more informed choices will optimize the use of the resources available and improve the efficiency of participatory breeding activities.

Keywords: *Zea mays* L., open-pollinated varieties, yield, nutritional quality, organoleptic quality, processing quality, genetic diversity, participatory plant breeding

1 Introduction

Maize (*Zea mays* L.) plays a major role in nutrition in many countries and is the basis for the production of several foods, such as polenta, bread, tortillas, snacks, and cornflakes (Fernandes et al., 2013). In some of the countries such as Spain or Portugal whole maize flour is used for bread production (Rodríguez et al., 2013). The ethnic Portuguese maize-based bread is known locally as *broa*. *Broa* is traditionally made with more than 50% maize flour mixed with rye and/or wheat flour in a mostly empirical process (Brites et al., 2010). As further described by the same authors (Brites et al., 2010), this process normally involves the mixing of sieved wholemeal maize flour

with hot water, rye and/or wheat flour (in a variable proportion), with yeast from leavened dough from earlier *broa* acting as sourdough.

In the last few decades, consumers' views on how foods positively or negatively affect their health have changed and, therefore, foods today are not only intended to satisfy hunger and provide necessary nutrients; they are also used to prevent nutrition-related diseases and improve physical and mental well-being (reviewed in Siró et al., 2008). Given this rising awareness in consumers, the consideration of the quality aspects of plant breeding is now a commercially relevant issue. The health benefits of consuming whole grains have been well documented, and are often associated with those benefits conveyed by their dietary fiber content (Ktenioudaki et al., 2015). Additionally, whole grains are rich in bioactive phytochemicals such as phenolic compounds, tocopherols, and carotenoids (Slavin et al., 2000).

Additionally, the market demand for gluten-free formulations has driven more research in the different steps leading from the maize kernel to the baking process (e.g., Moreira et al., 2015; Garzón et al., 2017; Martínez & Gómez, 2017). In parallel, an increased investment on the improvement of open-pollinated maize populations has been driven by a renewed interest in materials traditionally used for ethnic food commodities and for their use in the context of more sustainable farming systems (e.g., Revilla et al., 2012, 2015; Samayoa et al., 2016).

Since the introduction of maize in Europe from the Americas in the 15th century, diverse maize varieties have been selected for adaptation to a wide range of environments and consumer preferences (Revilla et al., 2015, Tenaillon & Charcosset, 2011). Portugal, Spain, and Italy are considered primary centers of maize introduction in Europe (Dubreuil et al., 2006). The European maize

populations although much less variable than the Central and South American populations (Rebourg et al., 2003), are a useful alternative because they were selected from multiple origins in the Americas and have the advantage of 400 years of adaptation to temperate climates (Romay et al., 2012), but lower yield than modern hybrids under conventional agricultural conditions (Revilla et al., 2015).

In the 21st century, Portuguese traditional maize populations can be still found under production as verified in a collecting expedition that took place in the last decade in the Central-Northern region of Portugal (Vaz Patto et al., 2007). This mission had as its main objective sampling the enduring traditional maize populations' variability in a particular region of the country, where maize-based bread still plays an important role in the local rural economy (Vaz Patto et al., 2007). In this collecting expedition, it was recorded that the majority of the maize populations conserved were being used primarily for bread production. As a consequence, the collected populations were assumed to have the potential to be used in *broa* production. The fact that flour produced from locally grown maize populations has traditionally been used in the formulation of *broa* has been pointed out by Vaz Patto et al. (2007) as one of the reasons for the on-farm conservation of the Portuguese maize populations.

Brites et al. (2010), through a sensory analysis on *broa* bread carried out by a trained panel using open-pollinated maize populations, identified a preference, due to texture, taste, and aroma, for maize bread produced using open-pollinated populations, as opposed to maize bread produced using commercial hybrid maize varieties. In the same study, instrumental quality attributes of maize flour from open-pollinated populations were measured and compared to commercial hybrid maize varieties. The results from that study showed that the flour from open-pollinated populations – considered

by the trained panel to produce better quality *broa* – had higher values of protein, lower values of amylose, and lower viscosities (maximum, minimum, final, and breakdown viscosities) (Brites et al., 2010).

Besides the phenotypic characterization, a better understanding of the genetic diversity present in the germplasm available for breeding helps to structure germplasm, defining, for example, heterotic pools; provides useful information for selecting contrasting parental lines for new breeding populations; and helps breeders to identify valuable new alleles for breeding (Varshney et al., 2016).

Currently, only a limited number of Portuguese traditional maize populations are integrated into the long-term participatory maize breeding program that has been running since 1984 in the northeast region of Portugal (Sousa Valley, Lousada) (Vaz Patto et al., 2013). One of the main advantages of on-farm participatory plant breeding is that it enables the constant adaptation of crops to the environment and supports the involvement of farmers since the selection criteria for the maize populations are defined in accordance with farmers' decisions. This breeding program was set at the Sousa Valley region because this was a well-known area in the country for maize production, with good edaphic-climatic conditions, and because at the time of the program implementation, it was initiated with the support of the local community (reviewed in Vaz Patto et al., 2013). In this Portuguese participatory maize breeding program, the selection was mainly focused on the improvement of grain yield and other important agronomic traits, considering that quality was safeguarded by the use of local traditional maize populations (Moreira, 2006). Nevertheless, by the comparative evaluation of different selection cycles of some of the participatory bred maize populations, Alves et al. (2017) verified that although diversity was maintained under this program, quality

traits evolved erratically. This observation, together with the increasing market importance given to quality aspects, set the stage for addressing the need to develop appropriate decision-making tools to bring about a quality-oriented maize population selection.

Although previous works (Vaz Patto et al., 2007, 2009; Brites et al., 2010) improved our knowledge of the agronomic, quality, and molecular aspects of traditional maize populations collected from the central region of Portugal, some limitations remained. Specifically, in terms of agronomic characterization, it is still necessary to understand the eventual effect and interaction of the different maize farming sites on those maize populations. Moreover, the use of controlled pollinations in the previous studies might have reduced production per plot, as described in Vaz Patto et al. (2007); therefore, field trials, under real production management over several locations, are still necessary to correctly evaluate the potential grain yield and to study how each traditional population behaves when grown in the different areas where maize populations have traditionally been produced in the country. In terms of quality characterization, it is necessary to evaluate other health-promoting, nutritional, and organoleptic compounds that can have an impact on consumers' perception and acceptance of the final product. Finally, in terms of molecular characterization, it is necessary to increase the number of individual plants evaluated per population from the original five, assessed in Vaz Patto et al. (2009). Maize is a naturally open-pollinated crop and, therefore, a large number of individuals should be evaluated to accurately estimate the number of alleles and their frequency per population and, as a result, to assess the similarities and infer the genetic structure between and within maize populations.

The maize populations that were surveyed in the collecting mission that took place in the Central-Northern region of Portugal

(Vaz Patto et al., 2007) are not at this date involved in any participatory maize breeding program. Given the previous Portuguese experience with this type of breeding approach and to promote the use of such distinct material, this work proposes to produce relevant (phenotypic and molecular) information on these materials and to develop decision-making tools to aid in the establishment of a quality-oriented participatory breeding program. This breeding program should take into consideration market-driven quality traits (traits related to consumer acceptance, such as organoleptic and health-related compounds), while also improving the agronomic performance of the breeding materials. The characterization of these populations will allow the identification of the most relevant ones for each breeding objective and will result in a more efficient use of those genetic resources in breeding programs. Therefore, the objectives of this study are:

(1) To extend the maize populations quality characterization – organoleptic, nutritional, and health-related traits – with the quantification of aroma-related volatile compounds, and health-related compounds, such as tocopherols, carotenoids, and phenolic compounds, that might influence the quality of maize-based food commodities;

(2) To accurately estimate the agronomic performance and potential of the collected maize populations using multi-location field trials (broader performance stability/specific adaptability) across different farming sites, exploring new locations for the establishment of a future quality-oriented participatory maize breeding program;

(3) To build decision-making tools to enable an accurate population selection within a quality-oriented participatory breeding program, by complementing the precise agronomic and quality description with a more thorough molecular characterization.

2 Materials and Methods

2.1 Plant material

The materials evaluated in this study consisted of 16 enduring traditional maize populations that were collected in the Central-Northern region of the country from small farms with low input agricultural systems (Vaz Patto et al., 2007). These farmers' populations were labeled in this work as *broa-x* (*x* corresponds to the specific name given to each population).

For comparison purposes, nine open-pollinated populations from the long-term Portuguese maize participatory breeding program, identified in this work as participatory bred (PPB) populations, and an international reference, the US open-pollinated population *BS22(R)C6*, were also included in this study. The populations under the Portuguese maize participatory breeding program were selected and/or developed primarily to improve their agronomic performance (reviewed in Vaz Patto et al., 2013). *BS22(R)C6* is a genetically broad-based synthetic population developed primarily for improved grain yield and root and stalk strength (Hallauer et al., 2000). More information about each population can be found in Table S1.

2.2 Quality evaluation

Quality traits related to flour's pasting behavior (flour viscosity parameters), nutritional value (protein, fat, and fiber content), bioactive compounds (carotenoids, tocopherols, total phenolic content, *p*-coumaric and ferulic acid content), and aroma-related compounds (volatile aldehydes content) were evaluated in 26 maize populations. For that, a bulk of kernel from each maize population produced from a common-garden experiment established in Coimbra in 2009 was used. Information about the site characterization can be

found in Table S2. Each population was overplanted by hand in two-row plots 6.4 m long and with 0.75 m border space between two planted rows. Each plot was thinned at the seven-leaf stage to 48 plants per plot to achieve a plant density of 50,000 plants.ha^{-1}. Plots were irrigated as needed and mechanically and/or hand weeded as necessary following common agricultural practices for maize in the region. Pollination was controlled within each plot. All the plots were harvested by hand. After harvest, ears were dried at 30-35ºC in an oven (Memmert Model UFE 800, Memmert GmbH + Co. KG, Germany) until a ~15% in moisture was reached. The ears were then shelled and the kernel collected per plot basis, packed in a paper bags and kept at 4ºC until further analysis.

Wholemeal maize flour was obtained after milling the kernel through a Cyclone Falling number 3100 mill (Perten, Sweden) with a 0.8 mm mesh.

The pasting properties of maize flour were obtained with a Rapid Viscosity Analyzer RVA-4 (Newport Scientific, Australia). The viscosity profiles were obtained for each population according to Almeida-Dominguez et al. (1997) at 15% solids, using the following heating and cooling cycle settings: (1) holding at 50°C for 2 min, (2) heating to 95°C in 4.5 min, (3) holding at 95°C for 4.5 min, (4) cooling to 50°C in 4 min, (5) holding at 50°C for 10 min. The RVA paddle speed was set at 960 rpm for the first 10 s of the test, after which the speed was changed to 160 rpm. Peak (PV), minimum or trough (TV), and final viscosities (FV) were recorded in cPoise and the breakdown viscosity (BD) was calculated as PV–TV, and setback from trough viscosity (SB1) was calculated as FV–TV.

Maize flour yellowness was determined on a 10 to 12 g sample in an opaque recipient using a Minolta chromameter CR-2b and the CIE tristimulus color parameters *b** (yellow/blue index). Positive *b**

values indicate that sample tends toward the yellow part of the color spectra.

Flour protein (PR), fat (FT), and fiber (FI) content were determined by a near-infrared spectroscopic method using Inframatic 8620 equipment (Perten, Sweden), with calibrations supplied by the manufacturer. Results were expressed in percentages.

The total carotenoids content (TCC) was spectrophotometrically measured at 450 nm according to the AACC method 14-60.01 (AACC International, 2012). Results were expressed in μgrams of lutein equivalent per gram of sample, as the main carotenoid found in maize.

α-Tocopherol (AT), γ-tocopherol (GT), δ-tocopherol (DT) were separated from the fat portion of the maize flours by high-performance liquid chromatography (HPLC) and quantified using an Agilent 1200 model with a fluorescence detector (FLD) and a Diol column (LiChropher 100, 250 x 4 mm) according to the method ISO 9936 (2006). Tocopherols content was expressed in μg/g fat basis.

For assessing the total free phenolic compounds content (PH) of maize flour ethanolic extracts (EtOH:H_2O 50:50, v/v) were prepared according to Lopez-Martinez et al. (2009), with some modifications. Briefly, 2 g of maize flour was extracted with 20 mL of EtOH:H_2O (50:50, v/v) for 15 minutes, using an Ultra Turrax T25 (Janke & Kunkel, IKA Labortechnik, Germany). Final extracts were filtered using a Whatman filter paper (type42: retention 2.5 μm, diameter 18.5 cm). Extracts were prepared in triplicate and preserved at -20ºC until analysis.

Total free phenolic compounds content (PH) was assessed using the *Folin-Ciocalteau* assay (Singleton et al., 1999) with a Beckman DU-70 spectrophotometer, with slight modifications as

described in Silva et al. (2015), and expressed in mg of gallic acid equivalents/100 g of dry weight (GAE/100 g DW).

p-Coumaric (CU) and ferulic acid (FE) were quantified by HPLC coupled with a photodiode array detector (HPLC-PDA) at 280 nm with a Thermo Finnigan Surveyor HPLC system according to Silva et al. (2006). *p*-Coumaric (CU) and ferulic acid content were expressed in mg/100 g of dry weight.

Solid phase micro-extraction (SPME) was used as sample preparation methodology and the volatile fraction was analyzed by gas chromatography-mass spectrometry (SPME-GC-MS). Briefly, to one gram of maize flour, 4.5 mL of Milli-Q water was added to a capped vial and were homogenized using a vortex. For sample preparation, a 2cm- 50/30 µm DVB/Carboxen/PDMS fiber (SUPELCO) and an exposure time of 60 minutes, at 60ºC were used.

Volatile compounds were analyzed in a GCMS-QP2010 Plus Shimadzu equipment and compounds were separated in a Varian Factor Four column (30 m x 0.25 mm x 0.25 µm). The injector was at 250ºC and the column was at 35ºC for 5 minutes, followed by a gradual increase of 5ºC/min until a final temperature of 230ºC was reached. The injection was performed using a splitless mode. The interface and ion source on MS equipment was set at 250ºC. Mass spectra were produced at 70 eV in a range of 29 – 299, using a scanning velocity of 555 scans/s. Helium was used as mobile phase at a flow rate of 2.1 mL/min. The equipment was coupled to an automatic sampler AOC-5000 (Shimadzu). GCMSsolution Release 2.53SU1 software was applied for data acquisition and treatment.

Volatile aldehydes content (AL) was taken as the sum of the peak area of the main aldehydes identified (hexanal, heptenal, 2-heptanal (*Z*), 2-octenal (*E*), nonanal, 2-nonenal (*E*) and decanal). Identification of volatile compounds was performed by a comparison

of the experimental mass spectra with the ones from the software's spectra library (WILEY 229, NIST 27 and 147). A standard mixture of hydrocarbons C8-C20 (40 mg/L each, in hexane) was used to determine linear retention indexes – LRI (Kovats indexes) – in order to confirm identification. The values of LRI determined for each compound were compared with described LRI for the same type of column (El-Sayed, 2014, http://www.pherobase.com).

2.3 Quality data analysis

All the calculations were performed in SAS Release 9.2 (SAS Institute Inc., 2004). Pearson correlation coefficients were calculated between the 14 maize quality traits in all maize populations using PROC CORR procedure.

Principal component analysis (PCA) was performed using the PROC PRINCOMP procedure on standardized data. The number of principal components was determined by checking eigenvalues of the principal components (Kaiser Criterion that retains components with eigenvalues greater than one and SCREE plot) and the cumulative proportion of variance explained.

The standardized principal component scores were multiplied by the root of their eigenvalues to calculate pairwise Euclidean distances between populations. The average linkage method (i.e., UPGMA) of PROC CLUSTER was applied in order to classify maize populations into groups and to determine the optimal number of clusters. Cubic Clustering Criterion (CCC) statistics and Pseudo F (PSF) statistics were calculated and plotted. The classification of maize populations into groups as obtained by cluster analysis was evaluated by discriminant analysis (DA) using 14 traits in PROC DISCRIM procedure in SAS. The probabilities of classification

success of the discriminant function were estimated by cross-validation.

The univariate analysis of variance using PROC GLM was conducted in order to test mean differences between quality-groups for 14 traits. Means were separated using the least-squares means procedure with Tukey's control adjustment for multiple comparisons.

2.4 Agronomic evaluation

The agronomic performance of all maize populations was compared in multi-location field trials. Field trials were established during 2010 in nine different sites: Quinta da Conraria, Montemor-o-Velho, S. Pedro do Sul, Lousada, Valada do Ribatejo, Vouzela-1, Vouzela-2, Travassos, and Coimbra.

The different locations represent different areas where maize open-pollinated populations traditionally are produced in the country and the different agronomic production systems normally associated with maize open-pollinated populations, ranging from conventional (Montemor-o-Velho) to organic (Quinta da Conraria and Valada do Ribatejo), and also considering low-input production systems (all the other locations). Information about the sites' characterizations can be found in Table S2.

During the 2010 growing season, a total of 26 maize populations were evaluated in a randomized complete block design, each population replicated within the three blocks set per field trial (location). Each population was overplanted by hand in two-row plots 6.4 m long and with 0.75 m between rows. Each plot was thinned at the seven-leaf stage to 48 plants per plot to achieve a plant density of 50,000 plants.ha^{-1}. Plots were irrigated as needed and mechanically and/or hand weeded as necessary. All the plots were harvested by hand.

In each environment, a maximum of 144 plants (48 plants per plot × 3 blocks) were evaluated for each population. Missing data issues were identified for all the late cycle populations (*Verdeal da Aperrela*, *Castro Verde*, *Estica*, *Fisga*, and *Fandango*) in Travassos, Vouzela-1, and S. Pedro do Sul; all sites located at mid-altitude, where no data was obtained. The *Pigarro* population, a participatory bred population, also suffered from poor adaptation to the trial environments since data for *Pigarro* could only be retrieved for three out of nine environments: Lousada (the population's site of origin), Valada do Ribatejo, and Vouzela-2, the latter with data in only one block.

Grain yield and ear weight per population were recorded for each block. Ear weight was taken as an indirect measurement of ear size, the trait for which the majority of the collected maize populations were being selected. The agronomic performance of each population was evaluated according to Moreira et al. (2008) as described in Table S3.

2.5 Agronomic data analysis

Pearson correlation coefficients between grain yield and ear weight were calculated using PROC CORR procedure in SAS Release 9.2 (SAS Institute Inc., 2004). Given the high correlation between grain yield and ear weight, further analysis on genotype by environment interactions was reported for grain yield only.

The genotype-by-environment (G × E) interaction analysis was carried out using Additive Main effects and Multiplication Interaction (AMMI) models, a convenient tool for detecting patterns and systemic trends that can usually have direct ecological or biological interpretation (Gauch et al., 2011). Previously described missing data issues required the model fitting using the Expectation-Maximization

(EM) algorithm, as implemented in the so-called “EM-AMMI” model (Gauch & Zobel, 1990).

The general form of AMMI models can be expressed as (Gauch, 1992):

$$Y_{ij} = \mu + g_i + e_j + \sum_{k=1}^{p} \lambda_k \gamma_{ik} \delta_{jk} + \rho_{ij} + \varepsilon_{ij}$$

where Y_{ij} is the mean response of the population i in the environment j; μ is the overall mean; g_i is the fixed effect of the population i (i = 1, 2, ... g); e_j is the fixed effect of environment j (j = 1, 2, ... e); ε_{ij} is the experimental error; the G × E interaction is represented by the factors λ_k , a singular value of the k^{th} interaction principal component axis (IPCA) (k = 1, 2, ... p, where p is the number of axes to be retained in the model), γ_{ik} , the population eigenvector for k^{th} IPCA, and δ_{jk} , the environmental eigenvector for k^{th} IPCA; ρ_{ij} is the residual comprised of the discarded axes.

Selection of the optimal model (number of axes to be retained in the model) was done by cross-validation, using two replicates for model fitting and the remaining one for validation in 1000 iterations. Both EM-AMMI modeling and cross-validation were carried out using MATMODEL software (Gauch, 2007).

After selecting the optimal AMMI model, the adaptability and phenotypic stability of the maize populations were summarized in a biplot. Since the optimal model was AMMI1, the biplot depicts the main effects of population/genotype and environment versus the scores for first IPCA. The biplot was generated in Microsoft Excel 2010 using the IPCA scores and trait means retrieved from MATMODEL software.

2.6 Molecular evaluation

Thirty random individual plants from each maize population were genotyped with 20 microsatellites (SSRs – simple sequence repeats). SSRs were chosen based on their location in the maize reference genome (1 SSR per chromosome arm), and repeat motifs (≥ 3 base pairs) to facilitate allele scoring (Table S4). Information about each SSR can be found at MaizeGDB (Lawrence et al., 2008 – www.maizegdb.org).

Genomic DNA was isolated from the adult leaves of each plant using the modified CTAB procedure as described in Saghai-Maroof et al. (1984). Genotyping procedures were carried out accordingly to Alves et al. (2017). A genotypic matrix of the alleles' scores per individual plant, in base pairs, was generated and served as the basis for the molecular data analysis.

2.7 Molecular data analysis

The informativeness of each microsatellite marker was assessed measuring their Polymorphism Information Content (PIC; Botstein et al., 1980) and the number of alleles detected using PowerMarker software (PowerMarker V3.23, Liu & Muse, 2005).

Genetic variability within each population was accessed by the following parameters: the average number of alleles per locus (N_{av}), the number of private alleles (N_{pr}), using GENEPOP software (GENEPOP V4.0, Raymond & Rousset, 1995), and the allelic richness (N_{ar}), as the measure of the number of alleles per locus independent of sample size, using FSTAT software (FSTAT V2.9.3.2, Goudet, 2002).

Also for each population, the following parameters based on the allelic frequencies were estimated: the observed (H_O) and expected heterozygosity (H_E), and the inbreeding coefficient (F_{IS}), using

GENEPOP software (GENEPOP V4.0, Raymond & Rousset, 1995). The same software was also used to test if the genotypic frequencies in each population were in conformance to Hardy-Weinberg (HW) expectations. The probability test for Hardy-Weinberg (HW) equilibrium was based on the Markov chain method (Guo & Thompson, 1992; Raymond & Rousset, 1995) followed by sequential Bonferroni adjustments (Rice, 1989) to correct for the effect of multiple tests, using SAS Release 9.2 (SAS Institute, 2004).

For comparison purposes, the significance of differences in average values of N_{ar}, H_O, H_E and F_{IS} between farmers' populations and participatory bred (PPB) populations were tested using FSTAT software (FSTAT V2.9.3.2, Goudet, 2002).

The genetic differentiation between all pairs of populations was measured with pairwise F_{ST} estimates. Pairwise F_{ST} values and their respective P-values for significant differences from zero were calculated with FSTAT software (FSTAT V2.9.3.2, Goudet, 2002).

To represent the genetic relationships between all maize populations, pairwise Cavalli-Sforza–Edwards' chord distances (D_{CSE}) (Cavalli-Sforza & Edwards, 1967) were calculated and an unrooted phylogenetic tree was constructed using Fitch-Margoliash algorithm (Fitch & Margoliash, 1967) with 1,000 bootstraps (Felsenstein, 1985) over microsatellite loci as implemented in SEQBOOT, GENDIST, FITCH, and CONSENSE programs of the PHYLIP software package (PHYLIP ver3.6b, Felsenstein, 2004).

The analysis of molecular variance (AMOVA, Excoffier et al., 1992) was used to partition the total microsatellite diversity among all populations and within all populations. The same analysis was also used to partition the total microsatellite diversity detected among farmers' populations and participatory bred populations, within farmers' populations *vs.* participatory bred populations, and within all

populations. The variance components retrieved from AMOVA analysis were used to calculate a series of statistics called φ-statistics, which summarize the degree of differentiation between population divisions and are analogous to Wright's F-statistics (Excoffier et al., 1992). The variance components were tested statistically by non-parametric randomization tests using 10,000 permutations in ARLEQUIN software (ARLEQUIN ver3.0, Excoffier et al., 2005).

A model-based clustering method was applied on multilocus microsatellite data to infer genetic structure and define the number of gene pools in the dataset using the STRUCTURE software (STRUCTURE V2.3.3, Pritchard et al., 2000). Given a value for the number of gene pools, this method assigns individual genotypes from the entire sample to gene pools in a way that linkage disequilibrium (LD) is maximally explained. Ten runs per each K were done by setting the number of gene pools (K) from 1 to 10. Each run consisted of a burn-in period of 200,000 steps followed by 10^6 MCMC (Monte Carlo Markov Chain) replicates assuming an admixture model and correlated allele frequencies. No prior information was used to define the gene pools. The choice of the most likely number of gene pools (K) was carried out by comparing the average estimates of the likelihood of the data, ln[Pr(X|K)], for each value of K (Pritchard et al., 2000), as well as by calculating an ad hoc statistic ΔK, based on the rate of change in the log probability of data between successive K values as described by Evanno et al. (2005). The program STRUCTURE HARVESTER was used to process the STRUCTURE results files (STRUCTURE HARVESTER v0.6.92, Earl, 2012).

3 Results

3.1 Quality evaluation

Correlations among quality traits can be found in Table S5. The majority (approximately 70%) of the quality traits were not correlated with each other, or had weaker correlations (46.34% of the total significant correlations detected), with a Pearson correlation coefficient $|r| < 0.5$. Protein (PR) content that was strongly positively correlated with fiber (FI) content ($r = 0.954$, $P < 0.001$). In addition, both these traits (PR and FI) were negatively correlated with the breakdown viscosity (BD) ($r = -0.752$ and $r = -0.711$, respectively, $P < 0.001$), and with the α-tocopherol ($r = -0.764$ and $r = -0.786$, respectively, $P < 0.001$) and δ-tocopherol values ($r = -0.693$ and $r = 0.719$, respectively, $P < 0.001$). The total carotenoids content (TCC) was strongly positively correlated with the flour yellowness ($r = 0.985$, $P < 0.001$), measure as *b** from the CIE tristimulus color parameters.

Because the parameters describing the pasting properties of maize flour were correlated among them, and because the breakdown viscosity (BD) and setback from trough viscosity (SB1) parameters were derived from the primary viscosity parameters (FV, PV, and TV), only the BD and SB1 viscosity parameters were chosen for further analyses.

A principal component analysis (PCA) on the standardized quality data was performed in order to summarize multivariate similarities among the maize populations analyzed.

The position of the maize populations along the first principal component (*x*-axis) in the PCA biplot, as shown in Figure 1, was mainly defined by their protein and fiber content, the breakdown viscosity, the total carotenoids content, α- and δ-tocopherol content,

and volatile aldehydes content. As shown in Figure 1, the farmers' populations (*broa-x* populations) were largely discriminated from the non-*broa-x* maize populations along this principal component. The position of the maize populations along the second principal component (*y*-axis) was set primarily according to its flour yellowness (measured by *b** color parameter), total carotenoids content, *p*-coumaric acid, and ferulic acid content. The third principal component was mainly influenced by setback from trough viscosity values, and the fourth principal component was mainly defined by the levels of total free phenolic compounds (Table S6).

To assess if the different maize populations under study would group into different quality-based groups, a cluster analysis was performed based on the first four principal components retrieved from the PCA. The first four principal components were used since we observed that only by considering the first four principal components, retrieved in the PCA, was a stabilized accumulated percentage of variance (77.94% of total variance) obtained, all having eigenvalues greater than one (Table S6).

As a result of the cluster analysis, the highest values of both Pseudo F (PSF) statistics and Cubic Clustering Criterion were obtained when considering three clusters. Therefore, it was decided that the classification of maize populations in three quality-groups would be the optimal solution. One of the clusters is composed exclusively of one population, the *Amiúdo* population, and was therefore excluded from further analyses. As for the other two quality-groups identified, one was mainly composed of farmers' populations (*broa-x* populations), and was named quality-group I; the second group identified was composed of the remaining maize populations and was named quality-group II (Figure 1).

The groups retrieved from cluster analysis were then validated by performing a discriminant analysis. The discriminant function, based on 14 traits, correctly classified all the populations into their respective quality-group (100% classification success) when using the standard method, and 22 out of 25 populations (88% classification success) when using the cross-validation method. The groups obtained by cluster analysis were in agreement with the populations' positions in the PCA biplot (Figure 1).

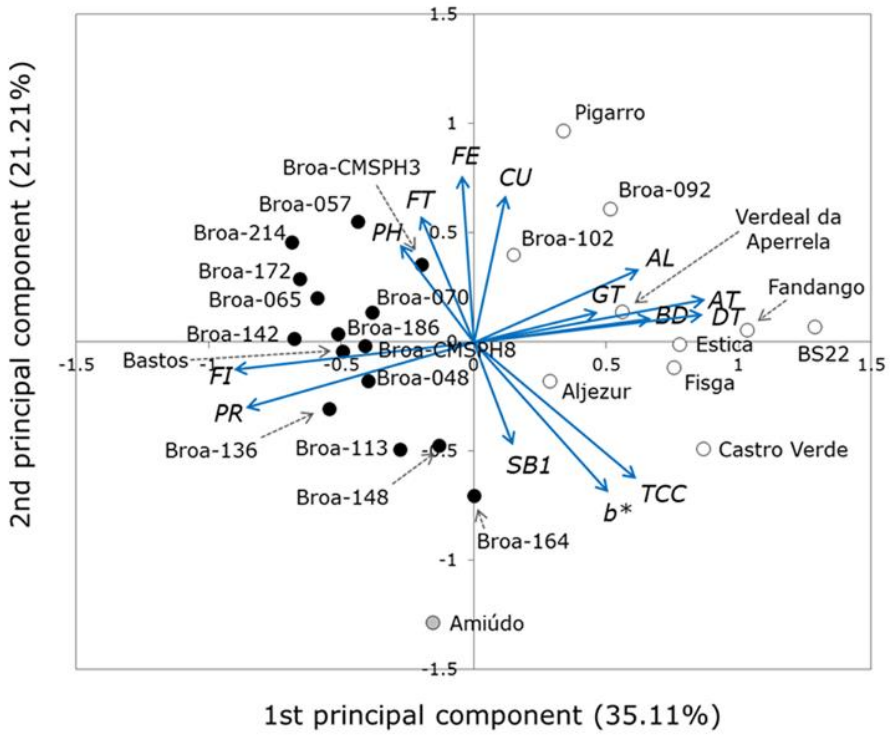

Figure 1. Biplot of principal component analysis (PCA) based on 14 quality traits measured in 26 maize populations; different colored circles correspond to the different quality-based groups identified on cluster analysis: quality-group I is depicted in black, quality-group II is depicted in white; the Amiúdo population is depicted in grey.

Quality-group I, where the majority of farmers' populations were clustered, was characterized by having a higher fiber and protein content than the average value found in quality-group II, and lower breakdown viscosity values, lower total carotenoids content, lower levels of volatile aldehydes, and lower α-tocopherol and δ-tocopherol content than the average values found in quality-group II (Table 1).

Table 1. Analysis of variance and comparison of mean values for the quality traits among quality-group I and quality-group II, as defined by cluster analysis.

Trait	Mean Square	P(F)[1]	Quality-group	
			I	II
Protein (PR)	31.89	***	12.18	9.83
Fiber (FI)	0.87	***	2.36	1.97
Fat (FT)	$1.47x10^{-5}$	ns	4.97	4.97
Breakdown (BD)	$2.54x10^{6}$	***	82.38	746.11
Setback1 (SB1)	$9.33x10^{6}$	ns	$1.97x10^{3}$	$2.37x10^{3}$
Yellow/blue index (*b**)	211.46	ns	16.72	22.78
Total carotenoids (TCC)	$2.31x10^{3}$	*	15.86	35.88
α-tocopherol (AT)	$20.07x10^{3}$	***	39.29	98.32
δ-tocopherol (DT)	627.43	***	16.21	26.65
γ-tocopherol (GT)	$8.49x10^{3}$	ns	244.26	282.65
Total free phenolic compounds (PH)	$1.08x10^{3}$	ns	159.64	145.92
p-coumaric acid (CU)	$5.48x10^{-3}$	ns	0.35	0.38
Ferulic acid (FE)	$4.48x10^{-4}$	ns	0.38	0.38
Volatile aldehydes (AL)	$6.84x10^{14}$	***	$2.44x10^{6}$	$13.34\ x10^{6}$

*[1]P(F) – significance of the F-test for differences between quality groups: ns – non-significant; * – significant at P < 0.05; *** – significant at P < 0.001*

Quality traits' units: Protein (PR), fiber (FI) and fat (FT) expressed in percentage; Viscosity parameters (BD and SB1) expressed in cPoise; Yellow/blue index (b) – if b* is positive it means that samples tend to the yellow part of the color spectra; Total carotenoids (TCC) expressed in µgrams of lutein equivalent per gram of sample; Tocopherols (AT, DT and GT) expressed in µg/g fat basis; Total free phenolic compounds content (PH) expressed in gallic acid equivalents/100 g of dry weight; p-coumaric acid (CU) and ferulic acid (FE) expressed in mg/100 g of dry weight; Aldehydes (AL) taken as the chromatogram peak area.*

3.2 Agronomic evaluation

Grain yield was strongly and positively correlated with ear weight ($r = 0.81$, $P < 0.0001$), therefore the following genotype-by-environment interaction analysis on agronomic data was reported only for grain yield.

The AMMI ANOVA (Table 2) shows that population, environment, and the G × E interaction were significant ($P < 0.05$) for grain yield. From the total variation expressed as the sum of squares, the genotypes accounted for 28.12%, and the G × E interaction accounted for a 16.96% variation. The cross-validation identified AMMI1 as the optimal model; therefore, G × E was further partitioned into a single interaction principal component axis (IPCA) and model residual.

Table 2. Additive Main effects and Multiplication Interaction (AMMI) analysis of variance for maize populations' grain yield tested in 9 different environments.

Source	Degrees of Freedom	Mean Square	P-value
Total	602	372.94	
Treatment	233	733.75	<0.001
Population	25	2525.58	<0.001
Environment	8	8719.55	<0.001
G × E [1]	200	190.34	<0.05
IPCA1[2]	32*	486.70	<0.001**
Residual	168	133.89	0.723
Error	369	145.11	

[1] G × E – Genotype-by-Environment interaction
[2] IPCA1 – first Interaction Principal Component Axis
** Degrees of freedom assigned to IPCAs using Gollob's method (Gauch, 1992)*
*** F ratio constructed using residual* mean square *as denominator*

The results of AMMI1 fitting for grain yield (Mg/ha) are illustrated in Figure 2. This biplot depicts both main effects for populations (G) and environments (E), on the *x*-axis, and G × E

interaction, on the *y*-axis. Coordinates, where the axes are crossing in the biplot, correspond to the overall grain yield mean (5.05 Mg/ha) (on the *x*-axis) and no G × E interaction (on the *y*-axis). The vertical axis separates lower-yielding populations and the environments where the maize populations performed the worst on the left side from the higher-yielding populations and environments where populations performed the best on the right side.

The population with the highest mean grain yield was *Fandango*, a participatory (PPB) bred maize population, and the population with the lowest mean grain yield was a farmers' maize population – *broa-142* (Figure 2). The horizontal axis separates all populations and environments into two groups with opposite interaction effects, and the strength of the interaction effects is depicted as the distance from the *x*-axis to each environment; therefore, the Coimbra site has the strongest positive interaction effect on the populations' performance and the Montemor-o-Velho site the strongest negative interaction effect on the populations' performance. The positioning of a population close to a certain environment indicates the specific adaptation of those populations to those environments.

Overall, all farmers' populations were low-yielding, with grain yield mean of 4.49 Mg/ha, value below the overall grain yield mean (5.05 Mg/ha), and with positive interaction effects with the Valada do Ribatejo, Travassos, and Coimbra sites; therefore, they are better adapted to those environments. Participatory bred populations with a long cycle until maturation (identified as late populations in Table S2), such as *Fandango*, *Estica*, *Fisga,* and *Verdeal da Aperrela*, had high grain yields (7.37 Mg/ha, 6.68 Mg/ha, 6.59 Mg/ha, and 5.85 Mg/ha, respectively) and performed better at environments such as the Montemor-o-Velho and Lousada sites.

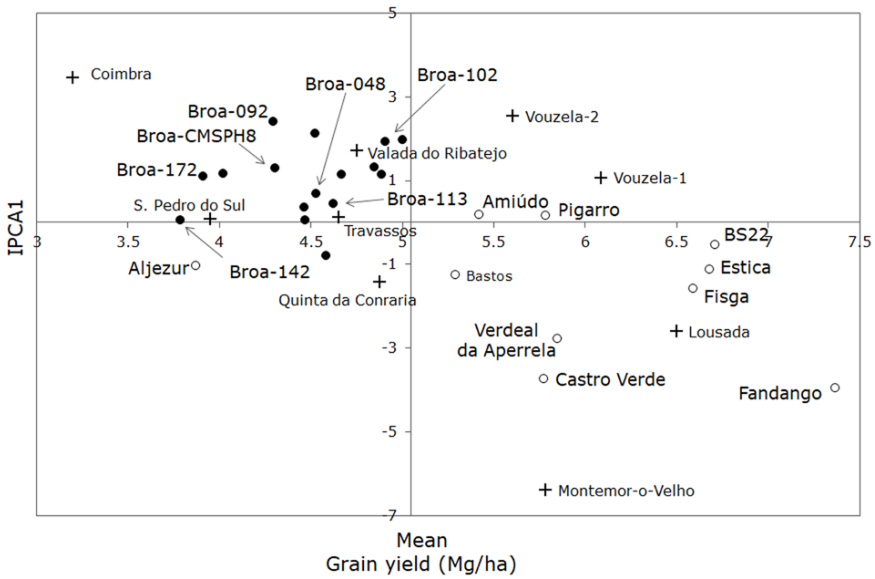

Figure 2. Biplot of mean grain yield against first principal component scores (IPCA1) of the Interaction Principal Component Analysis for 26 maize populations and 9 tested environments. Legend: farmers' populations are depicted in black circles; participatory bred (PPB) populations and the outer group (*BS22(R)C6*) are depicted in white circles; tested environments are depicted in black crosses.

3.3 Genetic diversity analysis

The molecular characterization of the populations was done using 20 microsatellites markers distributed evenly across the 10 maize chromosomes. The level of information retrieved from the markers used, calculated as the polymorphic information content (PIC), was, on average, 0.516. Overall the 20 microsatellites detected 114 different alleles, with an average of 5.7 alleles per marker (Table S4). Except for *broa-142*, from the farmers' populations, and *Verdeal da Aperrela*, from the participatory bred populations, both showing an excess of homozygous individuals (F_{IS} = 0.113 and F_{IS} = 0.093,

respectively), no deviations from Hardy-Weinberg expectations were detected in the remaining 24 maize populations (Table S7).

The results of the genetic variability assessment within each population can be found in Table S7. When considering only the farmers' populations (*broa-x* populations), the lowest number of alleles and the lowest genetic diversity (H_E) were found in population *broa-CMSPH8* (N_{ar} = 2.8; H_E = 0.405), whereas the highest values of both parameters were found in population *broa-113* (N_{ar} = 3.5; H_E = 0.549) (Table S7). For comparison purposes, it is worth noting that the US population (*BS22(R)C6*) always showed values of the number of alleles and genetic diversity below the average values detected on the farmers' populations (Table S7). It was also revealed that the allelic richness (N_{ar}) and genetic diversity (H_E) were significantly lower on farmers' populations when compared to participatory bred populations (N_{ar} = 3.164 *vs.* N_{ar} = 3.692; H_E = 0.490 *vs.* H_E = 0.514) (Table 3).

Genetic differentiation between all pairs of populations was measured with pairwise F_{ST} estimates. All pairwise F_{ST} values were significantly different from zero at $P < 0.05$, except between *Estica* and *Fisga* populations.

Table 3. Differences in average values of N_{ar}, H_O, H_E, and F_{IS} between farmers' populations and participatory bred (PPB) populations.

Group	No. of populations	N_{ar}	H_O	H_E	F_{IS}
Farmers' populations	16	3.164	0.487	0.490	0.008
PPB populations	9	3.692	0.514	0.544	0.055
P-value*		0.001	0.063	0.002	0.006

**P-values obtained after 1,000 permutations | N_{ar}: allelic richness; H_O: observed heterozygosity; H_E: expected heterozygosity; F_{IS}: inbreeding coefficient*

The average genetic differentiation of farmers' populations was below the overall average (overall F_{ST} = 0.124 *vs.* farmers' populations F_{ST} = 0.099) (Table S8).

The results from the analysis of molecular variance (AMOVA; Excoffier et al., 1992) can be found in Table 4. AMOVA was used to partition the total microsatellite diversity: (1) among and within all populations; (2) among farmers' populations and participatory bred populations, among populations within groups, and within all populations.

Table 4. Analysis of molecular variance (AMOVA) analysis for the partitioning of microsatellite diversity (1) among all populations and within populations, (2) among farmers' populations and participatory bred (PPB) populations, among populations within groups, and within all populations.

Analysis	Source of variation	df[1]	Percentage of variation	ϕ-statistics[2]	P-value (ϕ)[3]
(1) All populations	Among populations	25	12.75	ϕ_{ST} = 0.127	< 0.0001
	Within populations	1534	87.25		
(2) Farmers' populations *vs.* PPB populations	Among groups	1	2.30	ϕ_{CT} = 0.023	< 0.001
	Among populations within groups	23	10.29	ϕ_{SC} = 0.105	< 0.0001
	Within populations	1475	87.41	ϕ_{ST} = 0.126	< 0.0001

[1] *df* - stands for degrees of freedom | [2] *ϕ-statistics: corresponds to an analogous to the Wright's F-statistics which measures the degree of genetic differentiation* | [3] *P-value (ϕ): the level of significance of the ϕ-statistics was tested by non-parametric randomization tests using 10,000 permutations.*

The result from the AMOVA shows that most of the observed genetic variance (87.25%) can be explained by the heterogeneity that exists within each population – intra-population variability. Nevertheless, some degree of genetic differentiation exists between

farmers' populations and participatory bred populations with a ϕ_{CT} = 0.023 (P-value (ϕ) < 0.001) (Table 4).

In the unrooted tree, all farmers' populations were placed on the same branch, clustered together with two participatory bred populations – *Pigarro* and *Bastos*. Moreover, the farmers' populations were placed further away from the populations with a US genetic background – *BS22(R)C6*, *Fandango*, *Estica*, and *Fisga* (Figure 3).
The average genetic distance between all populations was 0.104, with the minimum distance observed between two participatory bred populations (*Estica* and *Fisga*, D_{CSE} = 0.021) and the maximum distance observed between a farmers' population – *broa-CMSPH8* – and the outer group population – *BS22(R)C6* – (D_{CSE} = 0.281) (Figure 3, Table S9).

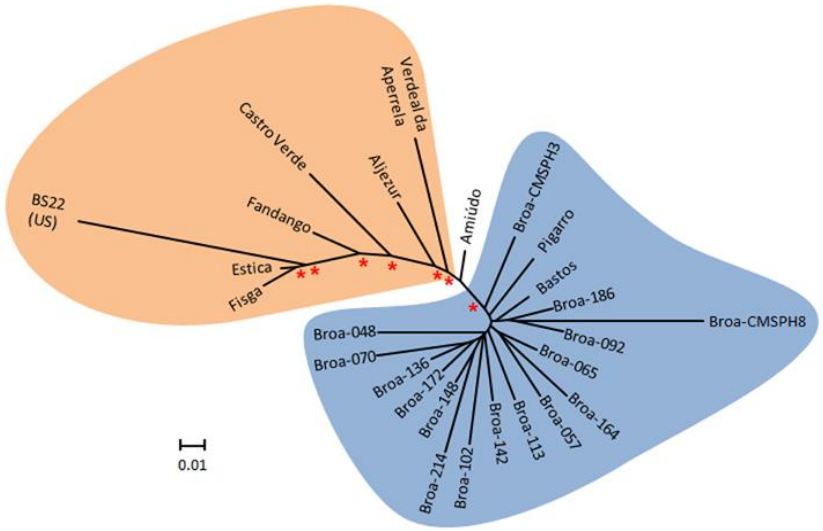

Figure 3. Fitch-Margoliash tree based on Cavalli-Sforza–Edwards' chord distances between 16 farmers' populations and 9 participatory bred (PPB) maize populations, plus the *BS22(R)C6* synthetic population from the US, abbreviated for *BS22* in the tree figure; bootstrap support values higher than 50% over 1,000 replicates are indicated with a red asterisk.

The existence of a genetic structure within the overall set of maize populations was investigated using a model-based clustering method implemented in STRUCTURE software (Pritchard et al. 2000). The highest ΔK value was observed for K = 2 (for K = 2, ΔK= 336.156, a value considerably bigger than the subsequent ΔK value for K = 3, ΔK = 67.031) and therefore two gene pools were considered to be the optimal solution. The proportion of membership of each gene pool in the 30 individual plants analyzed per population was retrieved from the run with the highest average estimates of the likelihood of the data, conditional on a given number of clusters, *ln*[Pr(X|K)].

From the 16 farmers' populations analyzed, all were predominantly build of gene pool A (Figure S1, gene pool A in blue), averaging a proportion of membership of 93.3% ± 9.6%.

4 Discussion

Given the previous successful Portuguese experience in participatory maize breeding and to promote the use of the maize populations collected from a *broa*-producing region, this work aimed to develop decision-making tools to support the establishment of a new participatory maize quality-oriented breeding program in the country.

4.1 Maize populations' quality characterization

The detailed characterization performed in the present study allowed for the identification of two main quality-based groups, and an outlying population, *Amiúdo*. *Amiúdo* clearly differed from the remaining maize populations in terms of its higher carotenoids level and lower levels of *p*-coumaric and ferulic acids. The different quality-based groups detected by cluster analysis were in agreement with the

results obtained from principal component analysis: 14 out of the 16 farmers' populations analyzed were placed in the same quality-group, named quality-group I, which corresponds to 87.5% of the farmers' populations (*broa-x* populations), with the exception of *broa-092* and *broa-102* populations; *broa-x* populations were essentially separated from the non-*broa-x* populations by their higher protein and fiber content, their lower levels of total carotenoids, α- and δ-tocopherol, and volatile aldehydes, as well as by their lower breakdown viscosity values. Populations belonging to quality-group I had on average 12.18% protein, a value slightly above the average reported for maize kernel (8–11% of protein, % w/w, FAO, 1992) but similar to the values (12.73–13.33%) previously reported by Vaz Patto et al. (2009) using an extended number of Portuguese maize populations. Quality-group I populations also presented on average 2.36% in fiber, which is similar to the value reported for maize kernel (2% fiber, % w/w, FAO, 1992; 2.59-2.61% in Vaz Patto et al., 2009). The populations from quality-group I had lower breakdown viscosities when compared with the populations from the other quality-group, which were composed mainly of non-*broa-x* populations. Breakdown viscosity (BD) is calculated as the difference between the peak (maximum) and the trough (minimum) viscosities obtained during the Rapid Visco Analyser (RVA) heating-cooling cycle. Breakdown viscosity is a measure of how easily the swollen starch granules can be disrupted after peak viscosity is reached during the RVA heating-cooling cycle (Wani et al., 2012). Since the breakdown viscosity is the result of the disintegration of starch granules, this value suggests the degree of starch stability during cooking (Wani et al., 2012). Julianti et al. (2015), when studying different composite flour formulations, observed that by increasing the proportion of soybean flour, a flour rich in protein, the breakdown viscosity measured during the RVA

heating-cooling cycle decreased. In the present work protein content and breakdown viscosity values are shown to have a strongly negative correlation between them. Related to what was discussed by Julianti et al. (2015), one of the possible explanations for the lower breakdown viscosities values observed in this current work in farmers' populations (*broa-x* populations) is the higher level of protein usually detected on those materials compared to the values obtained for the majority of non-*broa-x* populations.

It is known that the chemical composition of flour will influence the food texture and aroma (Collar et al., 2015; Shobha et al., 2015). Additionally, the maize populations that produce better-quality *broa* have higher protein values and lower breakdown values when compared to commercial maize varieties (Brites et al., 2010). The higher protein contents can probably induce increased amounts of flour water absorption ratio and corresponding higher bread moisture. In fact, the crumb moisture was been identified (Carbas et al., 2016) as a relevant attribute for consumer acceptability of *broa*.

Taking all that into consideration, according to the values of protein and breakdown viscosity obtained for traditional maize populations in the current work, and previously by Vaz Patto et al. (2009), one can argue that for maize populations used for *broa* production the optimal range values will be 12% to 13% of protein, and breakdown viscosity values of 82-300 cPoise.

Besides the basic nutritional value and pasting behavior-related traits also previously studied in Vaz Patto et al. (2009), in the current work, quality traits that might influence consumers' preferences/choices, such as volatile compounds related to aroma and health-related compounds such as carotenoids, tocopherols, and phenolic compounds, were also analyzed.

Vitamin A, as provitamin A carotenoids, and vitamin E, as tocopherols, are the predominant fat-soluble vitamins found in maize kernels (Nuss & Tanumihardjo, 2010). Moreover, the health benefits of grain products have also been associated with the antioxidant properties of the phenolic compounds found in grains (Bonoli et al., 2004). Carotenoids are a diverse family of yellow-orange pigments (Nuss & Tanumihardjo, 2010), and even though previous reports showed that grain color is not necessarily correlated with a provitamin A concentration of yellow and orange maize (e.g., Harjes et al., 2008), in the current work a strong positive correlation between the total carotenoid content and flour yellowness was detected.

Within the antioxidant phenolic compounds, ferulic acid is predominant in maize kernel, mainly present in the bound form (Adom & Liu, 2002), with *p*-coumaric acid also widely found in maize (Pei et al., 2016). Within the present study quality-group I, composed mainly by *broa-x* populations, a substantial range of variation could be found for flour yellowness and total carotenoids, and for the two individual phenolic compounds analyzed – *p*-coumaric acid and ferulic acid. This indicates that further improvement to increase the attractiveness of food formulations based on the populations within that quality-group, and specifically for those traits, where variation can still be found, is still possible. Indeed, some of these antioxidant compounds may reduce the retrogradation and improve starch qualities (Beta & Corke, 2004; Siriamornpun et al., 2016; Zhu et al., 2009), or influence the formation of dough texture (Klepacka & Fornal, 2006), a very important parameter in defining bread quality (Matos & Rosell, 2012).

Maize kernel nutritional composition can vary due to various factors such as the genotype, environmental conditions, and processing (Prasanthi et al., 2017). In the future, the study of G × E interaction for quality traits should also be undertaken since

genotype-by-environment interaction are known to affect some quality traits (e.g., Malvar et al., 2008; Revilla et al., 2015). This study would allow us not only to test the significance of the G × E on the presently considered quality traits but also to compare, for each trait, the proportion of explained variance by the G × E term with respect to the genotype main effects.

Because data acquisition for the quality traits accessed in this study is very expensive and time-consuming in the present work genotype-by-environment analysis was only performed at an agronomic level. Nevertheless, even with quality data from only one common-garden experiment, the results obtained from the multivariate analysis allowed us to highlight the similarities that exist among farmers' populations, as well as to identify the quality traits that discriminate them.

4.2 Maize populations' agronomic performance

Multi-location field trials were established across different farming systems in order to accurately estimate the agronomic performance and evaluate the agronomic potential of the farmers' maize populations. An Additive Main effects and Multiplicative Interaction (AMMI) method was implemented to identify maize populations with broader stability (i.e., lower variation across locations) or specific adaptability to the tested locations, and to evaluate potential new locations for the quality-oriented breeding program in the country. According to Furtado Ferreira et al. (2006), an undesirable population will have low stability associated with low productivity; therefore, the ideal population is one with high productivity and IPCA1 values close to zero (stable across environments).

The lower the IPCA1 value (in absolute values), the lower its contribution to the G × E interaction; therefore, the more stable the agronomic behavior of the population. On average, and in terms of grain production, the farmers' populations analyzed in the present work had a broader stability value when compared to all the maize populations ($|IPCA1|_{FARMERS} = 1.124$ *vs.* $|IPCA1|_{OVERALL} = 1.635$). However, the results also showed that all farmers' populations were low-yielding (4.49 Mg/ha, on average), performing better in environments such as the Valada do Ribatejo (organic production), Travassos, or Coimbra sites.

In conclusion, the agronomic evaluation allowed for the identification of the most appropriate locations where selection activities should be pursued if increasing grain yield and/or ear weight is among the breeding objectives in a quality-oriented participatory maize breeding program. Moreover, that choice can be fine-tuned according to the maize populations under selection. Of course, other factors, such as local support/interest from both farmers and local institutions (e.g., municipality and farmers' associations) must be taken into consideration when choosing the location for this kind of participatory research (Vaz Patto et al., 2013). In addition, the end product to be produced (maintaining the ethnic maize-based bread entity or extending it to other novelty food products) may influence the choice of the location as well as the particular populations that are more suitable due to their quality traits. In this way, if a population or a group of populations selected for a quality objective/end-use behaves better in a particular environment, this might be the best environmental choice. An extra factor to keep in mind for these decisions: should we consider the quality certification of the end product? For example, if we were to consider the Portuguese ethnic maize-based bread as a value-added product by adding a

certification, according to the European Union (EU) agricultural product quality policy (https://ec.europa.eu/agriculture/quality_en; accessed August 30th 2017), such as protected designations of origin (PDOs), protected geographical indications (PGIs), or traditional speciality guaranteed (TSG). This possibility of certification might have profound implications for the organization of the breeding program. Not only geographic implications (selection of the site(s) for PPB implementation), if one wants to select for a particular environment, but also on the breeding design/crosses allowed (intra-population selection, selection of one population *vs.* inter-population crosses, selection of several populations).

4.3 Phenotypic and molecular characterization data integration

One of the proposed objectives of this study was to build decision-making tools for an accurate population selection within a quality-oriented participatory breeding program. This was achieved by complementing a precise agronomic and quality description with a more thorough molecular characterization.

For example, in the case in which we need to start from either one particular population (intra-population selection) or from several populations (inter-population crosses), molecular information such as that gathered in this study acts as an effective extra decision-making tool to evaluate and compare the genetic resources available to breeders. As already pointed out by Reif et al. (2003), simple sequence repeat markers provide a valuable tool for grouping germplasm and are a good complement to field trials for identifying groups of genetically similar germplasm.

The genetic diversity/distance calculated between potential crossing parents can be chosen to assure the highest possible diversity within a cross (Tuvesson et al., 2007), to plan useful gene

combinations, increasing the performance through increased heterosis (Reif et al., 2003), or to add new variation to the breeding program in a controlled fashion (Tuvesson et al., 2007).

In the present work, based on the genetic distances and genetic structure of the maize populations, two main clusters could be identified that in a systematic manner separated the maize populations with a known US genetic background from the other maize populations. One of the clusters contained all the *broa-x* populations together with two participatory bred populations derived from two traditional maize populations (*Pigarro* and *Bastos*). The quality-group I, which is composed mainly of farmers' populations (14 *broa-x* populations), plus one participatory bred population (*Bastos*), is almost identical to this genetic-based cluster (only *Pigarro* is not included). We also observed that the maize flour from the majority of the *broa-x* Portuguese populations, evaluated at the Coimbra site, had higher levels of protein and fiber and lower levels of α- and δ-tocopherols, associated with a lower breakdown viscosity values when compared to the maize populations of quality-group II.

For illustration purposes, in the case of a quality oriented breeding program for maize bread using the Portuguese populations, one of the breeding objectives to be pursued could focus on increasing the agronomic performance of the populations and tocopherol levels (α- and δ-tocopherol content) that are limiting on this germplasm, but without compromising the protein content or increasing viscosity. An increase in maize vitamin E levels, as tocopherols, can elevate its nutritional value by enhancing their role as antioxidants (Nuss & Tanumihardjo, 2010). As an example, one can improve the α-tocopherol levels on these Portuguese populations by using as a donor parent the maize population with the highest α-tocopherol levels (*Fandango*; 123.64 µg/g fat basis; a population with

a known US genetic background). The cross with the *Fandango* population, genetically distant from the *broa-x* populations, may promote heterosis and consequently a higher agronomic performance of the resulting hybrid population.

As in the described example, the knowledge generated from both phenotypic and genotypic analysis will aid in deciding future breeding activities and genetic resources management. As for bread making and other end uses, the same decision-making process could be used to select the initial populations, breeding approaches, and optimal breeding locations. At present, existing information is already in use to identify potential maize open-pollinated populations as parental lines to generate better-performing population hybrids with increased content in tocopherols and total free phenolic compounds, decreased content in volatile aldehydes, and decreased overall viscosity. This information was compiled separately according to the populations' kernel color (white kernel *vs.* non-white kernel) since kernel color has been linked to consumer acceptance (Ranum et al., 2014) and also appears to be important for Portuguese maize bread consumer choices (Carbas et al., 2016).

Through the integration of the different levels of information available, more informed choices are optimizing the use of resources and improving the efficiency of participatory breeding activities.

Acknowledgments

MLA performed the DNA isolation and the SSR genotyping, the analysis of the molecular and quality data, and drafted the manuscript. CB and MP participated in the establishment of the field trials and performed the agronomic characterization. BC and MB performed the quality data acquisition. PMM coordinated the field trials and performed the agronomic data acquisition. CB and MRB

coordinated and participated in the acquisition of quality data, and participated in the manuscript revision. JG performed the analysis of the agronomic data, and critically participated in the manuscript revision. ZS participated in the analysis of the molecular and phenotypic data, and critically participated in the manuscript revision. MCVP designed and coordinated the study and participated in the drafting and revising of the manuscript.

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) and Horizon 2020 Framework Programme under grant agreements No 245058 and No 633571, respectively, and from the Fundação para a Ciência e a Tecnologia, through the research unit GREEN-IT (UID/Multi/04551/2013). MLA and MCVP were supported by Fundação para a Ciência e a Tecnologia (SFRH/BD/77723/2011; FCT Investigator 2015 contract - IF/01337/2014, respectively). The authors would also like to acknowledge Doctor Silas Pego and the farmers that were involved in the development of the maize populations, as well as to the Portuguese Bank of Plant Germplasm for providing the seeds from the participatory breeding populations used on this work. Original seed from the *BS22(R)C6* maize population was kindly supplied by Professor Arnel R. Hallauer.

References

AACC International (2012). Approved methods of analysis Method 14-60.01, 11th ed. Total carotenoid content of cereal grains and flours. St.Paul, MN, USA: AACC International.

Adom, K.K., & Liu, R.H. (2002). Antioxidant activity of grains. *Journal of Agricultural and Food Chemistry*, 50(21), 6182-6187.

Almeida-Dominguez, H. D., Suhendro, E. L., & Rooney, L. W. (1997). Factors affecting rapid visco analyser curves for the determination of maize kernel hardness. *Journal of Cereal Science*, 25(1), 93-102.

Alves, M.L., Belo, M., Carbas, B., Brites, C., Paulo, M., Mendes-Moreira, P., et al. (2017). Long-term on-farm participatory maize breeding by stratified mass selection retains molecular diversity while improving agronomic performance. *Evolutionary Applications*, 00:1–17.

Beta, T., & Corke, H. (2004). Effect of ferulic acid and catechin on sorghum and maize starch pasting properties. *Cereal Chemistry*, 81(3), 418-422.

Bonoli, M., Verardo, V., Marconi, E., & Caboni M.F. (2004). Antioxidant phenols in barley (*Hordeum vulgare* L.) flour: comparative spectrophotometric study among extraction methods of free and bound phenolic compounds. *Journal of Agricultural and Food Chemistry*, 52(16), 5195-5200.

Botstein, D., White, R.L., Skolnick, M., & Davis, R.W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics*, 32(3), 314–331.

Brites, C., Trigo, M.J., Santos, C., Collar, C., & Rosell, C.M. (2010). Maize-based gluten-free bread: influence of processing parameters on sensory and instrumental quality. *Food and Bioprocess Technology*, 3(5), 707-715.

Carbas, B., Vaz Patto, M.C., Bronze, M.R., Bento da Silva, A., Trigo, M.J., & Brites, C. (2016). Maize flour parameters that are related to the consumer perceived quality of 'broa' specialty bread. *Food Science and Technology*, 36, 259–267.

Cavalli-Sforza, L.L., & Edwards, A.W. (1967). Phylogenetic analysis: models and estimation procedures. *Evolution*, 21(3), 550-570.

Collar, C., Conte, P., Fadda, C., & Piga, A. (2015). Gluten-free dough-making of specialty breads: Significance of blended starches, flours and additives on dough behaviour. *Food Science and Technology International*, 21(7), 523-536.

Dubreuil, P., Warburton, M., Chastanet, M., Hoisington, D., & Charcosset, A. (2006). More on the introduction of temperate maize into Europe: large-scale bulk SSR genotyping and new historical elements. *Maydica*, 51, 281-291.

Earl, D.A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359-361.

El-Sayed, A. M. (2014). The Pherobase: Database of Pheromones and Semiochemicals. Accessed April 12, 2011. http://www.pherobase.com.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14(8), 2611-2620.

Excoffier, L., Laval, G., & Schneider, S. (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, 1, 47-50.

Excoffier, L., Smouse, P.E., & Quattro, J.M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131(2), 479-491.

FAO (1992). Maize in human nutrition. FAO Food and Nutrition Series Nº25, Rome, Italy.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 39(4), 783-791.

Felsenstein J. (2004). PHYLIP (Phylogeny Inference Package) Version 3.6. Distributed by the author. Department of Genomic Sciences, University of Washington, Seattle, WA.

Fernandes, T. J., Oliveira, M. B. P., & Mafra, I. (2013). Tracing transgenic maize as affected by breadmaking process and raw material for the production of a traditional maize bread, broa. *Food Chemistry*, 138(1), 687-692.

Fitch, W.M., & Margoliash, E. (1967). Construction of phylogenetic trees. *Science*, 155(3760), 279-284.

Furtado Ferreira, D., Garcia Borges Demétrio, C., John Manly, B.F., de Almeida Machado, A., & Vencovsky, R. (2006). Statistical models in agriculture: biometrical methods for evaluating phenotypic stability in plant breeding. *Cerne*, 12(4), 373-388.

Garzón, R., Rosell, C. M., Malvar, R. A., & Revilla, P. (2017). Diversity among maize populations from Spain and the United States for dough rheology and gluten-free breadmaking performance. *International Journal of Food Science & Technology*, 52(4), 1000-1008.

Gauch, H.G. (1992). Statistical analysis of regional yield trials: AMMI analysis of factorial designs. Amsterdam: Elsevier.

Gauch, H.G. (2007). MATMODEL version 3.0: Open source software for AMMI and related analyses. Crop and Soil Science, Cornell Univ., Ithaca, NY. Available at <http://www.css.cornell.edu/staff/gauch/matmodel.html>.

Gauch, H.G., Rodrigues, P.C., Munkvold, J.D., Heffner, E.L., & Sorrells, M. (2011). Two New Strategies for Detecting and Understanding QTL × Environment Interactions. *Crop Science*, 51(1), 96-113.

Gauch, H.G., & Zobel, R.W. (1990). Imputing missing yield trial data. *Theoretical and Applied Genetics*, 79(6), 753-761.

Goudet J. (2002). Fstat: A Program to Estimate and Test Gene Diversities and Fixation Indices. Version 2.9.3.2. Institut d'Ecologie, Bâtiment de Biologie, Université de Lausanne, CH-1015 Dorigny: Available at <https://www2.unil.ch/popgen/softwares/fstat.htm>

Guo, S.W., & Thompson, E.A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, 48(2), 361-372.

Hallauer, A.R., Russell, W.A., & White, P.R. (2000). Registration of BS21 (R) C7 and BS22 (R) C7 maize germplasm. *Crop Science*, 40(5), 1517-1517.

Harjes, C.E., Rocheford, T.R., Bai, L., Brutnell, T.P., Kandianis, C.B., Sowinski, S.G., et al. (2008). Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science*, 319(5861), 330-333.

ISO 9936 (2006). Animal and vegetable fats and oils – Determination of tocopherol and tocotrienol contents by high-performance liquid chromatography.

Julianti, E., Rusmarilin, H., & Yusraini E. (2015). Functional and rheological properties of composite flour from sweet potato, maize, soybean and xanthan gum. *Journal of the Saudi Society of Agricultural Sciences*, 16(2), 171-177.

Klepacka, J., & Fornal, Ł. (2006). Ferulic acid and its position among the phenolic compounds of wheat. *Critical Reviews in Food Science and Nutrition*, 46(8), 639-647.

Ktenioudaki, A., Alvarez-Jubete, L., & Gallagher, E. (2015). A review of the process-induced changes in the phytochemical content of cereal grains: the breadmaking process. *Critical Reviews in Food Science and Nutrition*, 55(5), 611-619.

Lawrence, C.J., Harper, L.C., Schaeffer, M.L., Sen, T.Z., Seigfried, T.E., & Campbell, D.A. (2008). MaizeGDB: The maize model organism database for basic, translational, and applied research. *International Journal of Plant Genomics*, 2008, 496957.

Liu, K., & Muse, S.V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, 21(9), 2128-2129.

Lopez-Martinez, L.X., Oliart-Ros, R.M., Valerio-Alfaro, G., Lee, C.H., Parkin, K.L., & Garcia, H.S. (2009). Antioxidant activity, phenolic compounds and anthocyanins content of eighteen strains of Mexican maize. *Food Science and Technology*, 42(6), 1187-1192.

Malvar, R. A., Revilla, P., Moreno-González, J., Butrón, A., Sotelo, J., & Ordás, A. (2008). White maize: genetics of quality and agronomic performance. *Crop Science*, 48(4), 1373-1381.

Martínez, M. M., & Gómez, M. (2017). Rheological and microstructural evolution of the most common gluten-free flours and starches during bread fermentation and baking. *Journal of Food Engineering*, 197, 78-86.

Matos, M.E., & Rosell, C.M. (2012). Relationship between instrumental parameters and sensory characteristics in gluten-free breads. *European Food Research and Technology*, 235(1), 107-117.

Moreira, P.M.R.M. (2006). Participatory maize breeding in Portugal. A case study. *Acta Agronomica Hungarica*, 54(4), 431-439.

Moreira, P. M. M., Pêgo, S. E., Patto, C. V., & Hallauer, A. R. (2008). Comparison of selection methods on 'Pigarro', a Portuguese improved maize population with fasciation expression. *Euphytica*, 163(3), 481-499.

Moreira, R., Chenlo, F., Arufe, S., & Rubinos, S. N. (2015). Physicochemical characterization of white, yellow and purple maize flours and rheological characterization of their doughs. *Journal of food science and technology*, 52(12), 7954-7963.

Nuss, E.T., & Tanumihardjo, S.A. (2010). Maize: a paramount staple crop in the context of global nutrition. *Comprehensive Reviews in Food Science and Food Safety*, 9(4), 417-436.

Pei, K., Ou, J., Huang, J., & Ou, S. (2016). *p*-Coumaric acid and its conjugates: dietary sources, pharmacokinetic properties and biological activities. *Journal of the Science of Food and Agriculture*, 96(9), 2952-2962.

Prasanthi, P. S., Naveena, N., Rao, M. V., & Bhaskarachary, K. (2017). Compositional variability of nutrients and phytochemicals in corn after processing. *Journal of food science and technology*, 54(5), 1080-1090.

Pritchard, J.K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945-959.

Ranum, P., Peña-Rosas, J.P., & Garcia-Casal, M.N. (2014). Global maize production, utilization, and consumption. *Annals of the New York Academy of Sciences*, 1312(1), 105-112.

Raymond, M., & Rousset, F. (1995). GENEPOP (version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity*, 86(3), 248–249.

Rebourg, C., Chastanet, M., Gouesnard, B., Welcker, C., Dubreuil, P., & Charcosset, A. (2003). Maize introduction into Europe: the history reviewed in the light of molecular data. *Theoretical and Applied Genetics*, 106(5), 895-903.

Reif, J.C., Melchinger, A.E., Xia, X.C., Warburton, M.L., Hoisington, D.A., Vasal, S.K., et al. (2003). Genetic distance based on simple sequence repeats and heterosis in tropical maize populations. *Crop Science*, 43(4), 1275-1282.

Revilla, P., Landa, A., Rodríguez, A., Ordás, A., & Malvar, R. A. (2012). Influence of growing and storage conditions on bakery quality of traditional maize varieties under organic agriculture. *Crop Science*, 52(2), 593-600.

Revilla, P., de Galarreta, J. I. R., Malvar, R. A., Landa, A., & Ordás, A. (2015). Breeding maize for traditional and organic agriculture. *Euphytica*, 205(1), 219-230.

Rice, W.R. (1989). Analyzing tables of statistical tests. *Evolution*, 43(1), 223-225.

Rodríguez, V. M., Soengas, P., Landa, A., Ordás, A., & Revilla, P. (2013). Effects of selection for color intensity on antioxidant capacity in maize (*Zea mays* L.). *Euphytica*, 193(3), 339-345.

Romay, M. C., Butrón, A., Ordás, A., Revilla, P., & Ordás, B. (2012). Effect of recurrent selection on the genetic structure of two broad-based Spanish maize populations. *Crop Science*, 52(4), 1493-1502.

Saghai-Maroof, M.A., Soliman, K.M., Jorgensen, R.A., & Allard, R.W. (1984). Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 81(24), 8014-8018.

Samayoa, L. F., Malvar, R. A., Moreno-González, J., Ordás, A., & Revilla, P. (2016). Evaluation of White Maize Populations for Quality and Agronomic Performance. *Crop Science*, 56(3), 1173-1178.

SAS Institute Inc (2004). The SAS system for windows, V8. 02. Cary, NC: SAS Institute.

Shobha, D., Vijayalakshmi, D., & Asha, K. J. (2015). Effect of maize based composite flour noodles on functional, sensory, nutritional and storage quality. *Journal of Food Science and Technology,* 52(12), 8032-8040.

Silva, S., Bronze, M.R., Figueira, M.E., Siwy, J., Mischak, H., Combet, E., et al. (2015). Impact of a 6-wk olive oil supplementation in healthy adults on urinary proteomic biomarkers of coronary artery disease, chronic kidney disease, and diabetes (types 1 and 2): A randomized, parallel, controlled, double-blind study. *The American Journal of Clinical Nutrition*, 101(1), 44-54.

Silva, S., Gomes, L., Leitão, F., Coelho, A.V., & Vilas Boas, L. (2006). Phenolic compounds and antioxidant activity of *Olea europaea* L. fruits and leaves. *Food Science and Technology International*, 12, 385-396.

Singleton, V.L., Orthofer, R., & Lamuela-Raventos, R.M. (1999). Analysis of total phenols and other oxidant substrates and antioxidants by means of Folin-Ciocalteu reagent. *Methods in Enzymology*, 299, 157-178.

Siriamornpun, S., Tangkhawanit, E., & N. Kaewseejan. (2016). Reducing retrogradation and lipid oxidation of normal and glutinous rice flours by adding mango peel powder. *Food Chemistry*, 201, 160-167.

Siró, I., Kápolna, E., Kápolna, B., & Lugasi, A. (2008). Functional food. Product development, marketing and consumer acceptance—A review. *Appetite*, 51(3), 456-467.

Slavin, J.L., Jacobs, D., & Marquart, L. (2000). Grain processing and nutrition. *Critical Reviews in Food Science and Nutrition* 40(4), 309-326.

Tenaillon, M.I., & Charcosset, A. (2011). A European perspective on maize history. *Comptes Rendus Biologies*, 334(3), 221-228.

Tuvesson, S., Dayteg, C., Hagberg, P., Manninen, O., Tanhuanpää, P., Tenhola-Roininen, T., et al. (2007). Molecular markers and doubled haploids in European plant breeding programmes. *Euphytica*, 158(3), 305-312.

Varshney, R.K., Singh, V.K., Hickey, J.M., Xun, X., Marshall, D.F., Wang, J., et al. (2016). Analytical and decision support tools for genomics-assisted breeding. *Trends in Plant Science*, 21(4), 354-363.

Vaz Patto, M.C., Alves, M.L., N.F., Almeida, C.S., Mendes, P., & Satovic, Z. (2009). Is the bread making technological ability of Portuguese traditional maize landraces associated with their genetic diversity? *Maydica*, 54(2/3), 297-311.

Vaz Patto, M.C., Mendes-Moreira, P.M., Alves, M.L., Mecha, E., Brites, C., Bronze, R., et al. (2013). Participatory plant quality breeding: An ancient art revisited by knowledge sharing. The Portuguese experience. In *Plant breeding from laboratories to fields,* S. B. Andersen (Ed.), (pp. 255–288). Rijeka,Croatia: InTech.

Vaz Patto, M.C., Moreira, P.M., Carvalho, V., & Pego, S. (2007). Collecting maize (*Zea mays* L. convar. *mays*) with potential technological ability for bread making in Portugal. *Genetic Resources and Crop Evolution*, 54(7), 1555-1563.

Wani, A.A., Singh, P., Shah, M.A., Schweiggert-Weisz, U., Gul, K., & I. A. Wani. (2012). Rice starch diversity: Effects on structural, morphological, thermal, and physicochemical properties—A review. *Comprehensive Reviews in Food Science and Food Safety*, 11,417-436.

Zhu, F., Cai, Y.Z., Sun, M., & Corke, H. (2009). Effect of phytochemical extracts on the pasting, thermal, and gelling properties of wheat starch. *Food Chemistry*, 112(4), 919-923.

Supplementary material

Tables

Table S1. Maize populations under study, seed source, information on previous selection efforts, brief morphological characterization of the populations (endosperm type, grain color and population maturation), populations' geographic origin and geographic coordinates of the populations from the collection mission (Vaz Patto et al., 2007).

Source	Population	Population Type	Farmers' selection	Selection objective	Endosperm type	Grain color	Maturation	Geographic origin	Geographic Coordinates (Latitude / Longitude)	Altitude (masl[3])
Farmer	Broa-048	Traditional	Yes	Wide and longer ears	flint	yellow	early	Central Portugal	40º50'32"N / 7º56'14"W	471
Farmer	Broa-057	Traditional	Yes	Best and bigger ears	flint	white	early	Central Portugal	40º54'49"N / 7º58'22"W	502
Farmer	Broa-065	Traditional	NA	NA	flint	white	early	Central Portugal	40º57'05"N / 7º54'52"W	876
Farmer	Broa-070	Traditional	Yes	Best ears full of grains	flint	white	early	Central Portugal	40º41'48"N / 8º04' 58"W	567
Farmer	Broa-092	Traditional	Yes	Best ears	flint	white	early	Central Portugal	40º38'02"N / 8º03'42"W	429
Farmer	Broa-102	Traditional	Yes	Earliness, smaller plants, smaller grain and full ears	flint	yellow	early	Central Portugal	40º39'55"N / 7º59'00"W	424
Farmer	Broa-113	Traditional	Not currently	Bigger and fascinated ears	flint	orange	early	Central Portugal	40º45'04"N / 7º34'26"W	609

Continuation of Table S1

Source	Population	Population Type	Farmers' selection	Selection objective	Endosperm type	Grain color	Maturation	Geographic origin	Geographic Coordinates (Latitude / Longitude)	Altitude (masl[3])
Farmer	Broa-136	Traditional	Not currently	Bigger ears	flint	white	early	Central Portugal	40°19'30"N / 7°41'10"W	795
Farmer	Broa-142	Traditional	Not currently	Fasciated ears	flint	white	early	Central Portugal	40°28'33"N / 7°29'03"W	801
Farmer	Broa-148	Traditional	NA[2]	NA	flint	yellow	early	Central Portugal	40°31'31"N / 7°34'13"W	459
Farmer	Broa-164	Traditional	Yes	Wider ears	flint	orange	early	Central Portugal	40°39'01"N / 7°24'32"W	423
Farmer	Broa-172	Traditional	NA	NA	flint	white	early	Central Portugal	40°19'55"N / 7°50'33"W	269
Farmer	Broa-186	Traditional	Yes	Smaller plants and bigger ears	flint	white	early	Central Portugal	40°25'19"N / 7°55'19"W	338
Farmer	Broa-214	Traditional	Yes	Earliness, bigger ears with no diseases or lodging	flint	white	early	Central Portugal	40°49'43"N / 8°13'06"W	660
Farmer	Broa-CMSPH3	Traditional	NA	NA	flint	white	early	Central Portugal	40°16'13.74"N / 8°16'49.97"W	130
Farmer	Broa-CMSPH8	Traditional	NA	NA	flint	white	early	Central Portugal	41°07'1.75"N / 8°05'12.75"W	321
PPB[1]	Amiúdo	Traditional		Yield	flint	yellow	early	Northern Portugal		
PPB	Bastos	Traditional		NA	flint	white	early	Northern Portugal		

Continuation of Table S1

Source	Population	Population Type	Farmers' selection	Selection objective	Endosperm type	Grain color	Maturation	Geographic origin	Geographic Coordinates (Latitude / Longitude)	Altitude (masl[3])
PPB	Pigarro	Traditional		Bigger years	flint	white	early	Northern Portugal		
PPB	Verdeal da Aperrela	Traditional		NA	flint	white	late	Northern Portugal		
PPB	Aljezur	Traditional		NA	flint	yellow	early	Southern Portugal		
PPB	Castro Verde	Traditional		Bigger years	flint	yellow	late	Southern Portugal		
PPB	Estica	Synthetic		Longer years	dent	yellow	Late	80% US, 20% Portugal		
PPB	Fisga	Synthetic		Prolificacy	dent	yellow	Late	80% US, 20% Portugal		
PPB	Fandango	Synthetic		Yield	dent	yellow	Late	80% US, 20% Portugal		
USA	BS22(R)C6	Synthetic		yield, root and stalk strength	dent	yellow	early	US		

[1] PPB stands for Portuguese participatory maize breeding program | [2] NA stands for information not available | [3] masl stands for meters above sea level

Table S2. Location, soil and climate characterization of the field trials sites. – Table available online through the link <https://figshare.com/s/8e6803ff1cb901c2aab6>

Table S3. List of agronomic traits evaluated per plot basis, abbreviation and respective description.

Trait	Abbreviation	Units/Scale	Description
Ear weight	EW	Gram (g)	Ear weight, adjusted to 15% of grain moisture. Measure by averaging the weight of 4 shelled ears per plot.
Grain yield[1]	Y	Kilogram/hectare (kg/ha)	Grain yield adjusted to 15% moisture. Formula: Grain yield = Ear weight × (Grain weight/Ear weight) × (100%–% moisture at harvest)/ (100%–15% moisture). Grain weight and ear weight taken from 4 shelled ears.

[1]*Grain yield adjusted to 15% of moisture was calculated according to Moreira et al. (2008)*

Table S4. List of 20 microsatellite loci, their repeat motifs, chromosomal bin positions, and allelic diversity within 26 maize populations (N = 780).

No.	Marker	Repeat motif	Bin location	Range (bp)	N_a[1]	PIC[2]
1	nc007	(CCT)	5.01	143 – 158	6	0.701
2	phi059	(ACC)	10.02	142 – 160	6	0.416
3	phi065	CACTT	9.03	131 – 156	5	0.501
4	phi084	(GAA)	10.03-10.04	148 – 160	4	0.442
5	umc1065	(ACA)17	2.06	128 – 164	13	0.832
6	umc1134	(AGC)	7.03	79 – 91	7	0.491
7	umc1139	(GAC)4	8.01	146 – 158	3	0.283
8	umc1267	(CGG)4	9.03-9.04	113 – 122	4	0.582
9	umc1329	(GCC)7	4.06	77 – 86	3	0.426
10	umc1425	(TCA)4	3.04	119 – 131	5	0.368
11	umc1431	(GCA)5	1.09	132 – 138	3	0.54
12	umc1689	(GCG)5	1.05	137 – 146	4	0.403
13	umc1690	(GCA)4	3.07	82 – 94	4	0.425
14	umc1777	(CTG)4	8.05	113 – 125	5	0.412

Continuation Table S4

No.	Marker	Repeat motif	Bin location	Range (bp)	N_a[1]	PIC[2]
15	umc1787	(CGG)4	7.02	80 – 89	3	0.411
16	umc2030	(CGA)4	2.04	112 – 124	5	0.574
17	umc2059	(CAG)8	6.08	122 – 146	8	0.727
18	umc2196	(CCG)	6.01	115 – 133	6	0.552
19	umc2216	(TC)10	5.06	118 – 134	9	0.492
20	umc2281	(GTCC)5	4.03	152 – 204	11	0.734
	Average				5.7	0.516
	Total				114	

[1] *N_a – total number of alleles*
[2] *PIC – Polymorphism Information Content*

Table S5. Pearson correlation coefficients among 17 quality traits calculated from a common-garden experiment (Coimbra, 2009) for 26 maize populations analyzed. – Table available online through the link <https://figshare.com/s/4dc7d834959ecdf37125>

Table S6. Pearson correlation coefficients between quality traits and the first four principal components (PC) scores (PC1 to PC4), and the eigenvalues and percentage of variance for the four principal components.

No.	Trait	PC1		PC2		PC3		PC4	
1	PR	-0.863	***	-0.300	ns	0.326	ns	-0.095	ns
2	FI	-0.907	***	-0.124	ns	0.222	ns	-0.121	ns
3	FT	-0.203	ns	0.580	**	-0.402	*	-0.020	ns
4	BD	0.669	***	0.107	ns	-0.582	**	-0.018	ns
5	SB1	0.147	ns	-0.472	*	-0.717	***	-0.155	ns
6	*b**	0.506	**	-0.688	***	0.223	ns	0.029	ns
7	TCC	0.611	***	-0.626	***	0.250	ns	0.014	ns
8	AT	0.872	***	0.195	ns	0.039	ns	-0.118	ns
9	DT	0.863	***	0.128	ns	0.120	ns	-0.176	ns
10	GT	0.468	*	0.139	ns	0.533	**	-0.553	**
11	PH	-0.281	ns	0.447	*	-0.210	ns	-0.718	***
12	CU	0.120	ns	0.676	***	0.132	ns	0.552	**
13	FE	-0.046	ns	0.765	***	0.265	ns	-0.045	ns
14	AL	0.624	***	0.336	ns	0.289	ns	0.094	ns
Eigenvalue		4.915		2.969		1.796		1.231	
% of variance		35.11		21.21		12.83		8.79	

*P-value of the significance levels of correlations indicated as: ns – non-significant; * – significant at P < 0.05; ** – significant at P < 0.01; *** – significant at P < 0.001*

Quality traits' abbreviations: PR – protein; FI – fiber; FT – fat; BD – breakdown; SB1 –setback1; b* *– yellow/blue index; TCC – total carotenoids; AT – α-tocopherol; DT – δ-tocopherol; GT – γ-tocopherol; PH – total free phenolic compounds; CU –* p-*coumaric acid; FE – ferulic acid; AL – volatile aldehydes*.

Table S7. Within-population genetic diversity estimates in 26 maize populations (N = 780).

Population	N	N_{av}	N_{ar}	N_{pr}	H_O	H_E	F_{IS}	P-value HWE
Broa-048	30	3.2	3.1	1	0.502	0.468	-0.073	ns
Broa-057	30	3.4	3.3	0	0.462	0.477	0.032	ns
Broa-065	30	3.5	3.4	1	0.487	0.507	0.041	ns
Broa-070	30	3.2	3.1	0	0.487	0.504	0.033	ns
Broa-092	30	3.3	3.2	0	0.458	0.482	0.049	ns
Broa-102	30	3.6	3.5	0	0.466	0.490	0.049	ns
Broa-113	30	3.6	3.5	0	0.566	0.549	-0.032	ns
Broa-136	30	3.3	3.2	0	0.514	0.525	0.020	ns
Broa-142	30	3.3	3.2	1	0.474	0.535	0.113	**
Broa-148	30	3.0	2.9	0	0.488	0.483	-0.010	ns
Broa-164	30	3.1	3.0	0	0.491	0.463	-0.059	ns
Broa-172	30	3.4	3.3	1	0.506	0.508	0.004	ns
Broa-186	30	3.2	3.1	2	0.490	0.487	-0.007	ns
Broa-214	30	3.0	2.9	0	0.446	0.445	-0.003	ns
Broa-CMSPH3	30	3.1	3.0	0	0.525	0.518	-0.015	ns
Broa-CMSPH8	30	2.8	2.8	1	0.424	0.405	-0.048	ns
Amiúdo	30	4.0	3.8	1	0.503	0.526	0.042	ns
Bastos	30	3.5	3.4	0	0.503	0.530	0.052	ns
Pigarro	30	3.8	3.7	0	0.493	0.523	0.057	ns
Verdeal da Aperrela	30	3.5	3.4	1	0.468	0.516	0.093	***
Aljezur	30	4.0	3.9	0	0.566	0.591	0.042	ns
Castro Verde	30	3.8	3.7	0	0.457	0.498	0.082	ns
Estica	30	4.1	4.0	0	0.549	0.588	0.065	ns
Fisga	30	3.9	3.8	1	0.536	0.560	0.043	ns
Fandango	30	3.7	3.6	0	0.551	0.563	0.022	ns
BS22(R)C6	30	2.9	2.8	0	0.477	0.468	-0.019	ns
Average		3.4	3.3		0.496	0.508	0.022	

*N – sample size; N_{av} – average number of alleles; N_{ar} – average number of alleles per locus independent of sample size (allelic richness); N_{pr} – total number of private alleles; H_O – observed heterozygosity; H_E – expected heterozygosity; F_{IS} – inbreeding coefficient; P-value HWE: The probability global test for Hardy-Weinberg equilibrium (HWE) for each population was based on Markov chain method. ns – non-significant; ** – significant at P < 0.01; *** – significant at P < 0.001*

Table S8. Pairwise F_{ST} values between farmers' populations, pairwise F_{ST} values between participatory bred (PPB) populations, and pairwise F_{ST} values between all maize populations.

Parameter	Farmers' populations		PPB populations		All populations	
	F_{ST}	Between populations	F_{ST}	Between populations	F_{ST}	Between populations
Average	0.099		0.113		0.124	
Minimum	0.030	Broa-142 / Broa-172	0.005	Estica / Fisga	0.005	Estica / Fisga
Maximum	0.262	Broa-214 / Broa-CMSPH8	0.233	Verdeal da Aperrela / Castro Verde	0.377	Broa-CMSPH8 / BS22(R)C6

F_{ST} stands for fixation index

Table S9. Average chord distance between farmers' populations, average chord distance between participatory bred (PPB) populations, and overall Cavalli-Sforza–Edwards' chord distances between all maize populations.

Parameter	Farmers' populations		PPB populations		All populations	
	D_{CSE}	Between populations	D_{CSE}	Between populations	D_{CSE}	Between populations
Average	0.078		0.096		0.104	
Minimum	0.035	Broa-136 / Broa-172	0.021	Estica / Fisga	0.021	Estica / Fisga
Maximum	0.183	Broa-214 / Broa-CMSPH8	0.164	Verdeal da Aperrela / Castro Verde	0.281	Broa-CMSPH8 / BS22(R)C6

D_{CSE} - Pairwise Cavalli-Sforza–Edwards' chord distance

Figures

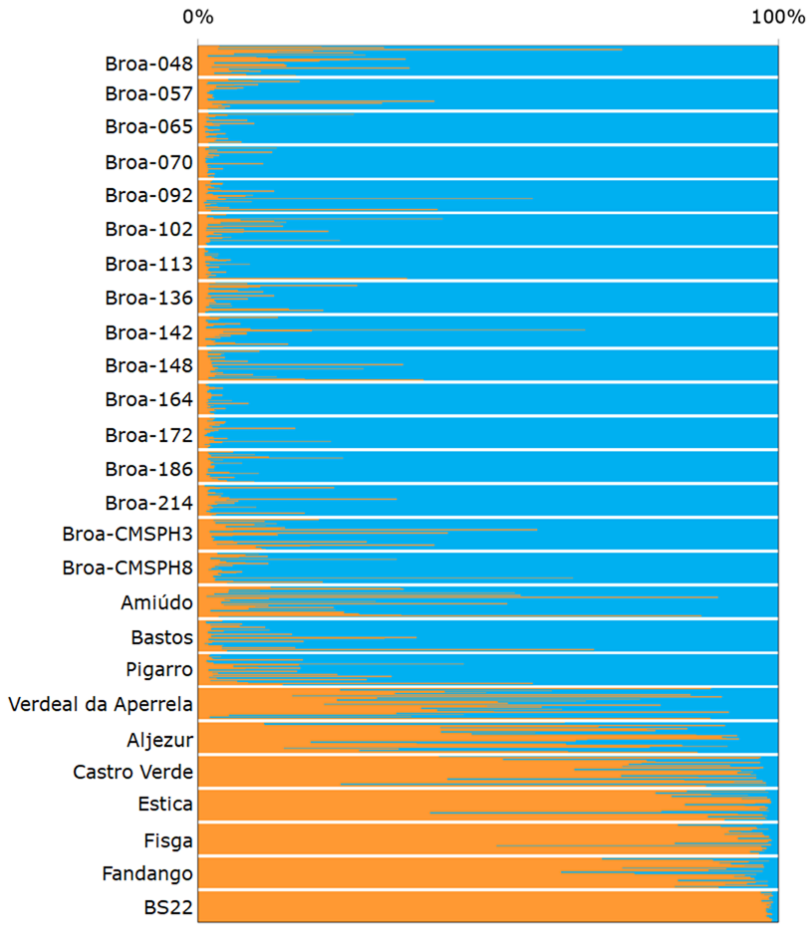

Figure S1. Proportion of membership of each maize population in each of the two gene pools inferred from multi-locus microsatellite data using a model-based clustering method. Each horizontal line within each population corresponds to an individual plant. Gene pool A is depicted in blue; gene pool B is depicted in orange. The *BS22(R)C6* synthetic population from the US, abbreviated for *BS22* in the figure.

Chapter IV

Genome-wide association study for kernel composition and flour pasting behavior in wholemeal maize flour

The work presented in this chapter corresponds to the following manuscript in preparation:

Alves M. L., Carbas B., Gaspar D., Paulo M., Brites C., Mendes-Moreira P., Brites C., Malosetti M., van Eeuwijk F., & Vaz Patto M. C. Genome-wide association study for kernel composition and flour pasting behavior in wholemeal maize flour (*in preparation*)

In this research paper, Mara Lisa Alves performed the DNA isolation, the genotypic and phenotypic data analysis, the association mapping analysis and follow-up analysis, and drafted the manuscript. (See Acknowledgements section for authors' contributions)

Abstract

Maize is a crop in high demand for food purposes and consumers worldwide are increasingly concerned with food quality. However, breeding for improved quality is a complex task and therefore developing tools to select for better quality products is of great importance nowadays. In Portugal, a unique germplasm has been developed through centuries of adaptation to local environment and food uses, in particular, for ethnic maize leaved maize *broa* bread production. Several parameters related to kernel composition, flour pasting behavior and flour particle size have been previously identified as crucial for *broa* quality.

In this work we took advantage, for the first time, of an original collection of 132 maize inbred lines, partially developed from Portuguese traditional maize populations, and carried a genome-wide association study aiming to identify genomic regions controlling compositional and pasting properties of maize wholemeal flour, fundamental on the development of quality-related molecular selection tools. The inbred lines were trialed during two growing seasons and samples from each field replicate characterized for main compositional traits (protein, fiber, fat, and starch), flour pasting parameters (viscosity profiles) and mean particle size, using well-established methodologies. The same collection was genotyped with the MaizeSNP50 BeadChip array. SNP-trait associations were tested using a mixed linear model that accounted for the genetic relatedness among inbred lines.

With this approach, 57 genomic regions were identified associated with the 11 different quality traits evaluated. Several regions controlling multiple traits were also detected with the identification of potential candidate genes. As an example, for

breakdown viscosity and peak viscosity, two viscosity parameters that reflect the capacity of the starch to absorb water and swell, the strongest common associated region for both traits was located near the *dull endosperm 1* gene which encodes a starch synthase and is determinant on the starch endosperm structure in maize.

This study allowed a better understanding of the complex genetic basis of maize kernel main compositional and pasting quality, identifying candidate genes for the majority of the quality associated genomic regions, or the most promising target regions to develop molecular tools to increase efficacy and efficiency of quality selection within maize breeding programs.

Keywords: *Zea mays* L., nutritional quality, pasting behavior, Portuguese maize germplasm, candidate genes, plant breeding

1 Introduction

Maize (*Zea mays* L.) is, along with rice and wheat, one of the world leading crops and a crucial source of food, feed, fuel and fibers (Tenaillon & Charcosset, 2011). Together, these three species account for 93% of all cereal food consumption (FAO, 2012). The major nutritional components and source of economic value in maize kernel are starch, protein, oil, and fiber (reviewed in Chen et al., 2016). From a processing perspective, the maize kernel is composed of four primary structures. The endosperm, germ, pericarp, and tip cap make up 83%, 11%, 5%, and 1% of the maize kernel biomass, respectively (Gwirtz & Garcia-Casal, 2014). The endosperm is primarily composed of starch surrounded by a protein matrix (Gwirtz & Garcia-Casal, 2014). The germ or embryo is high in fat (33.3%), as well as in enzymes and nutrients needed for maize plant growth and development (Gwirtz & Garcia-Casal, 2014). In the maize kernel

endosperm, starch is deposited as semi-crystalline granules, constituted by amylose and amylopectin. Although starch metabolism is complex, only four classes of enzymes, adenosine diphosphate glucose pyrophosphorylases, starch synthases, starch branching enzymes and debranching enzymes, have been identified as critical players in starch biosynthesis (Wang et al., 2015, and references therein). The primary storage proteins in the maize kernel are prolamines, called zeins, which are divided into four subfamilies of α- (19 and 22-kDa), β- (15 kDa), γ- (16-, 27-, and 50-kDa), and δ-zeins (10- and 18-kDa) (reviewed in Hartings et al., 2012).

The way in which maize is processed and consumed varies greatly from country to country, with maize flour and meal being two of the most popular products (Ranum et al., 2014, and references therein). A better understanding of the complex genetic basis of maize kernel main components is essential for devising more efficient breeding tools to support the improvement of this crop main products compositional quality.

Maize mutants have been widely used to isolate genes encoding key enzymes in starch metabolism (e.g., *Shrunken1* (*sh1*), *Brittle2* (*bt2*), *Waxy1* (*wx1*), *Sugary2* (s*u2*), *Dull1* (*du1*), and *Sugary1* (*su1*) (reviewed in Wang et al., 2015), as well as genes regulating zein synthesis and deposition (e.g., *opaque-2* (*o2*), *opaque-15* (*o15*), *floury1* (*fl1*)) (reviewed in Hartings et al., 2012). High-throughput genomics and post-genomics approaches are now providing new tools to better understand the genetic and biochemical networks operating during maize kernel development, contributing ultimately to its composition, and a high degree of complexity and regulation has been detected (reviewed in Hartings et al., 2012).

Quantitative trait loci (QTL) linkage mapping, and for the last 15 years, association mapping studies enabling higher resolution QTL

location, have shown that kernel main components are controlled by many genes, having complex patterns of inheritance (e.g., Cook et al., 2012, and references therein). For instance, Wilson et al. (2004) used an association approach to evaluate the involvement of six maize candidate genes, from the starch biosynthesis pathway, on major kernel compositional related traits (protein, oil, and starch concentration and composition, including pasting properties and amylose levels). With this work, Wilson et al. (2004) identified haplotypes of *brittle endosperm2* (*bt2*), *shrunken1* (*sh1*), and s*hrunken2* (*sh2*) that were associated with several kernel composition traits, and haplotypes of *amylose extender1* (*ae1*) and *sh2* with association with starch pasting properties (Wilson et al., 2004). More recently, genome-wide association studies (GWAS) have been used on the genetic dissection of maize kernel quality traits. Examples are the work of Li et al. (2013) that carried out a GWAS to unravel the genetic architecture of oil biosynthesis in maize kernels, and the work conducted by Cook et al. (2012) that carried out a joint-QTL mapping/GWAS for kernel starch, protein, and oil content.

The Portuguese maize germplasm is recognized by its high diversity (Vaz Patto et al., 2004, 2007) and associated potential quality for food since Portugal has a long tradition in the production of the ethnic leavened maize-based bread – *broa* (Vaz Patto et al., 2007). This ethnic bread is made with a 50% or more of maize flour, mixed with wheat or rye (Brites et al., 2010), for which the local maize populations are usually preferred (Vaz Patto et al., 2007). Several of the maize flour parameters that mainly influence maize kernel quality for *broa* production have been identified (Brites et al., 2010; Carbas et al., 2016). Protein and amylose content, flour pasting parameters, such as maximum, minimum and final viscosities (Brites et al., 2010), and flour particle size (Carbas et al., 2016), were among these major

influencing parameters. Pasting properties of maize flour are considered important parameters to consider for the preparation of different food products as they are related to its swelling and gelatinization characteristics (Paraginski et al., 2014). Starch, proteins, and lipids are the three major food components in cereal-based food products, and interactions among them in a food system are of importance to functionality and quality (Wang et al., 2017; Zhang & Hamaker, 2003).

A comprehensive analysis of all these different quality-related parameters is still missing in the Portuguese maize germplasm and so the national diversity was never properly exploited on the development of efficient tools / innovative approaches to support breeding for these complex quality traits.

The present study was carried out to identify genomic regions controlling the upper mentioned quality-related parameters through a genome-wide association approach, using a unique association panel constituted by a collection of maize inbred lines where a considerable amount of the unexplored Portuguese maize germplasm is present. This will allow the understanding of the genetic architecture of quality traits, the identification of candidate quality genes and the development of quality associated molecular selection tools for traits difficult to select by conventional methods. In this work, we took advantage of the diverse germplasm developed through decades of maize breeding by the extinct NUMI (Núcleo de Melhoramento de Milho), and now conserved at the Portuguese Plant Germplasm Bank (Banco Português de Germoplasma Vegetal - BPGV, Braga, Portugal). The uniqueness of the association mapping panel used in the current work, constituted by Portuguese, foreign and mixed origin lines, could lead to the discovery of quality alleles not previously identified in other germplasm collections.

Our approach consisted of (i) phenotyping the germplasm collection with different quality parameters, using samples harvested from a two environments field experiment, (ii) genotyping the same germplasm collection with the MaizeSNP50 BeadChip array, (iii) investigating the degree of genetic structure within the collection, and (iv) performing a GWAS with a mixed linear model approach, with the subsequent search for candidate genes and /or associated molecular markers.

2 Materials and Methods

2.1 Plant material

The maize inbred line collection used on this study was assembled observing a significant representation of lines selected from traditional Portuguese maize populations (29 lines) and lines with a mixed Portuguese x foreign origin (the majority of the lines whose names start by *PB*, *PP*, *PV* or *PT*, Table S1). The reasoning behind this was the premise that the locally grown Portuguese maize populations, is the material traditionally used for the formulation of baked commodities (as the leaved maize-based bread *broa*), are considered as keepers of quality traits related to bread production. The original seed of the maize inbred lines collection used in this study was provided by the Portuguese Bank of Plant Germplasm (BPGV, Braga, Portugal).

From a total of 164 different maize inbred lines sowed on the field trials, only 132 yielded sufficient kernels to proceed with their quality analysis (Table S1). Additional details on their recorded pedigree may be found in Table S1. Thirty-six of the yielding lines had a white kernel, further divided into 20 with flint endosperm, three intermediate and 13 with dent endosperm. The remaining 96 inbred

lines had a kernel color ranging from yellow to red, further divided into 37 with flint endosperm, eight intermediate, and 51 with dent endosperm (Table 1).

Table 1. Summary of the number of maize inbred lines grouped accordingly to their kernel color and endosperm type used to measure nutritional and processing quality traits in wholemeal flour.

Kernel color	Endosperm type			
	Flint	Intermediate	Dent	Total
White	20	3	13	36
Yellow	3	4	26	33
Yellow-orange	18	3	23	44
Orange	16	1	1	18
Red	−	−	1	1
Total	57	11	64	132

2.2 Field characterization and experimental design

The inbred lines were evaluated at Coimbra site (40°13'0.22"N, 8°26'47.69"W), Portugal, during the 2011 and 2012 growing seasons, using an organic agriculture converted field. The conversion started in 2011 and the field was considered to be fully managed under an organic agriculture system by 2012. This site is part of the Mondego river irrigation perimeter, a very high-yielding maize area where the average maize hybrids yield is 14.5 $Mg.ha^{-1}$ (Mendes-Moreira et al., 2015). It is located 50 km from the seacoast, with 25 m altitude. Its alluvial soils are characterized at 0-20 cm and 20-40 cm respectively by a pH of 5.65 and 5.75, a percentage of soil with a particle size less than 0.2 mm diameter of 83.37% and 82.84%; and an organic matter percentage of 2.91% and 2.55%. Agricultural practices were similar in both growing seasons, but sowing and harvest dates differed between growing seasons. Sowing took place at the 28th April and 11th May and the harvests took place on the 28th September and on the 6th November in 2011 and 2012, respectively.

In each year, the maize inbred lines were evaluated using a randomized complete block design, with two blocks (replicates). Information on the spatial distribution of the plots was also recorded (row and columns field coordinates). Each plot consisted of two rows 7.2 m long (6.4 m planted row plus 0.8 m border space between two planted rows), with an inter-row distance of 0.75 m. Each plot was overplanted by hand and thinned at the V7 growth developmental stage to achieve a plant density of approximately 50,000 plants ha^{-1}. Plots were mechanically and hand-weeded when needed and managed following common agricultural practices for maize in the region. Pollination was controlled within each plot. All the plots were harvested by hand. After harvest, ears were dried at 30-35ºC in an oven (Memmert Model UFE 800, Memmert GmbH + Co. KG, Germany) until a ~15% in moisture was reached. The ears were then shelled, and the kernel collected per plot basis packed in paper bags and kept at 4ºC until further analysis.

2.3 Phenotypic data acquisition

A seed sample from each of the harvested plots (replicates) was used for quality determinations. Therefore, the total number of samples analyzed corresponded to [*number of inbred lines* × *number of field replicates (2)* × *number of growing seasons (2)*].

Wholemeal maize flour was obtained from all the seed samples using a Falling number 3100 mill (Perten Inc., Sweden) with a 0.8 mm screen.

In order to prevent/minimize the enzymatic action and subsequent alteration of the flour properties, flour samples were also lyophilized using Cientificolab® equipment built for pilot-scale lyophilization of food commodities. For that, each sample was

individually placed in a flask (height 3.7 cm, diameter 4.2 cm) and then freeze-dried for long-term preservation.

Eleven quality traits were measured in wholemeal maize flour: nutritional-related traits – protein (PR), fiber (FI), fat (FT) and total starch content (ST and STL, see below for details); technological-related traits – mean particle size (SIZE and SIZE L; see below for details), peak (or maximum) viscosity (PV), trough (or minimum) viscosity (TV), final viscosity, breakdown viscosity (BD), setback from trough viscosity (SB1), and setback from peak viscosity (SB2).

2.3.1 Flour protein, fiber, fat and starch content

Flour protein (PR), fiber (FI) and fat (FT) content were determined for each non-lyophilized sample by near-infrared reflectance (NIR) spectroscopy (Percon Inframatic 8620, Perten Inc., Sweden), with calibrations for non-lyophilized samples supplied by the manufacturer. Values for protein, fiber, and fat corresponded to the mean value of up to two technical replicates. The total starch content was determined in lyophilized (STL) (2011 and 2012 growing seasons) and non-lyophilized (ST) (only 2012 growing season) samples using Fourier Transform Near-Infrared Reflectance (FT-NIR) spectroscopy (FT-NIR MPA, Bruker Optics, Germany), with calibrations for non-lyophilized samples supplied by the manufacturer. Values for total starch content obtained from 2012 growing season lyophilized and non-lyophilized samples where further used to test whether both datasets were correlated (phenotypic correlation between datasets). Values for total starch content (non-lyophilized (ST) and lyophilized (STL) samples) corresponded to the mean value of two to four technical replicates. Protein, fiber, fat, and starch content was expressed in percentage (%).

2.3.2 Mean particle size

The maize flour Particle Size Index (PSI) was determined using also FT-NIR spectroscopy (FT-NIR MPA, Bruker Optics, Germany). For 2011 growing season, only the mean for particle size in lyophilized samples (SIZEL) was measured. For 2012 growing season, both mean particle size in non-lyophilized (SIZE) and lyophilized flours (SIZEL) were determined. Values for mean particle size (non-lyophilized (SIZE) and lyophilized (SIZEL) samples) corresponded to the mean value of two to four technical replicates. The calibration models for PSI FT-NIR analysis were obtained using the particle size values measured in a subset of 30 non-lyophilized samples according to the AACC method 55-40.01:1999 (AACC, 1999), with a Malvern multi-channel laser light-scatter instrument (Malvern Instruments Ltd., England). Values for mean particle size obtained from lyophilized and non-lyophilized samples from the 2012 growing season were further used to test whether both datasets were correlated (phenotypic correlation between datasets). After calibration, the mean particle size volume value, or D[4,3], retrieved from the particle size distribution, was used as an average measure of the particle size of each sample and was expressed in µmeters.

2.3.3 Flour pasting properties

Maize flour pasting properties were evaluated by recording their viscosity profiles using a Rapid Visco Analyser (RVA) (Newport Scientific, Australia). The viscosity profiles were obtained on non-lyophilized samples according to Almeida-Dominguez et al. (1997) at 15% solids, using the following heating and cooling cycle set: (1) holding at 50°C for 2 min, (2) heating to 95°C in 4.5 min, (3) holding at 95°C for 4.5 min, (4) cooling to 50°C in 4 min, (5) holding at 50°C for 10 min. The RVA paddle speed was set at 960 rpm for the first 10

s of the test, after which the speed was changed to 160 rpm. Peak (or maximum) (PV), trough (or minimum) (TV) and final (FV) viscosities were recorded. The breakdown (BD) was calculated as peak viscosity-trough viscosity, setback from trough viscosity (SB1) as final viscosity - trough viscosity, and setback from peak viscosity (SB2) as final viscosity - peak viscosity. Up to two technical replicates of the viscosity profiles were taken for each sample. All the viscosity and viscosity-related traits were expressed in cPoise.

2.4 Phenotypic data analysis

A phenotypic analysis was performed per individual trial to 1) perform quality control of the data, 2) obtain estimates of genetic variances (and covariances between traits) and heritability, and 3) obtain adjusted trait means per inbred line. For quality control, graphical inspection of residuals was used to assess normality (Q-Q plot), homogeneity of variance (residuals versus fitted values), and identify outliers. Potential influential observations identified by the raw data method, which identifies observations exceeding 1.5 times the interquartile range, and the residual method, which identified standardized residuals by mixed model analysis, were removed from the analysis. One of the traits (breakdown viscosity, BD) required a squared-root-transformation to stabilize the variance. All analyses were done using the Breeding View software (Murray et al., 2014), available through the IBP Breeding Management System (The IBP Breeding Management System Version 3.0.9, 2015).

In detail, single trait-single growing season analysis, using mixed models, was performed using the “Single trait field trial analysis” pipeline of Breeding View, selecting the model for resolvable row-by-column design as implemented in the software. The statistical model includes an intercept, a fixed block effect, a

random row and column effects (nested within blocks), a genotypic effect (fixed or random, see the explanation that follows) and a residual. The Field trial analysis node in Breeding View performs two mixed model analyses: in the first step (Step 1) the inbred lines (genotypes) were fitted as a random term, while in the second step (Step 2) the inbred lines were fitted as a fixed term. The Step 1 model is used to obtain estimates of variance parameters. From Step 1 the heritability, as well as the best linear unbiased predictors (BLUPs) were calculated for each inbred line (and correlations between BLUPs of different traits used to obtain estimates of genetic correlations between traits). In Step 2, structural variance components (rows and column variances) are fixed to those estimated in Step 1, and by including the inbred lines as a fixed term, best linear unbiased estimators (BLUEs) for each inbred line were produced.

For each quality trait, a multi-environment trial analysis was also performed to assess the consistency across growing seasons. The analysis of variance was carried out using the REML variance components analysis procedure in Genstat software (Genstat® for Windows 18th edition, Payne et al., 2015). The mixed model included growing seasons (fixed), maize inbred lines and season by line interaction (fixed or random) while blocks, rows, and columns, were treated as random terms, and nested within growing seasons. Similarly to what was already described for the single trial analysis, in the multi-environment trial analysis, BLUPs and BLUEs were calculated for each inbred line across growing seasons. BLUPs were used on principal component analysis (PCA) to assess genetic correlations between traits and BLUEs were used as input phenotypic data in the association mapping analysis, for the combined analysis across growing seasons.

2.5 Genotypic data acquisition

DNA was isolated from adult leaves from each maize inbred line using a modified CTAB procedure as described in Saghai-Maroof et al., (1984). DNA quality was accessed using a 0.8% SeaKem® LE Agarose gel (Cambrex Bio Science Rockland, Inc., USA) stained with SYBR® Safe (Invitrogen, USA). DNA quantification was done using a spectrophotometer Nanodrop ND-2000C (Thermo Scientific, USA). An additional step for polysaccharides removal (Rether et al. 1993) was added when the ratio 260/230 nm wavelength was inferior to 1.6 to avoid the interference of these contaminants on Single Nucleotide Polymorphism (SNP) genotyping. DNA concentration for all inbred lines was set to 50 ng/µl and genotyped with the Illumina MaizeSNP50 BeadChip array (Ganal et al., 2011). The genotyping array procedure and alleles scoring was conducted by the genotypic service provider (TraitGenetics GmbH, Gatersleben, Germany). This array allows the screening of 17,520 genes (since 33,417 of the SNPs present in this array are located on 17,520 genes and 16,168 SNPs are located in intergenic regions) (Ganal et al., 2011). The position of each marker along the maize B73 reference genome was updated from the initially available coordinates when the MaizeSNP50 BeadChip was originally designed (B73 reference genome version 1) to the coordinates in the released B73 reference genome version 3. These coordinates were taken from the maize genome browser, via the MaizeGDB database (Lawrence et al., 2008, www.maizegdb.org).

2.6 Genotypic data analysis

2.6.1 Genotypic data quality control

Genotypic data quality control was performed by removing SNP markers and inbred lines with more than 25% of missing data. SNPs called as heterozygous were set as missing data (0.93% of the total SNP calls). Moreover, markers close to fixation (allelic frequency superior to 95%) or markers with a minor allele frequency (MAF) smaller than 5% were also removed. After this filter, a total of 48,772 SNPs remained and were used for the association mapping analysis.

2.6.2 Genetic structure analysis

A subset of 1,821 SNPs, evenly distributed across the genome (corresponding approximately to 1 SNP per Megabase pairs, Mb), was used to calculate principal components to study the population structure among inbred lines and to calculate the kinship matrix to study the pairwise genetic relatedness among inbred lines as implemented in Genstat software (Genstat® for Windows 18th edition, Payne et al., 2015).

2.7 Association mapping analysis

Given that for all the quality traits under study, the variance components for genotype-by-environment (G × E) interaction ($\sigma^2_{g\times y}$) were much smaller than the genotype variance component (σ^2_g), univariate association analysis was carried out using the adjusted means for field trial design (BLUEs) obtained across growing seasons. Genome-wide association studies were conducted with the Genstat software using the available genotypic (SNPs scored with the MaizeSNP50 BeadChip array) and quality data (11 quality traits) measured in 132 maize inbred lines. The Genstat software performs association mapping in the mixed model framework, fitting markers as

fixed and inbred lines as random terms using REML (Malosetti et al., 2007).

Three different models were tested to detect significant marker-trait associations: the naïve model [Phenotype = SNP + (Genotype + Error)], that neither accounts for population structure nor familiar relatedness; a model accounting for population structure (Q) using 15 principal components from PCA [Phenotype = Q + SNP + (Genotype + Error)]; and a model accounting for familiar relatedness (K) [Phenotype = SNP + Genotype + Error] with Genotype random effects structured following a kinship matrix K. For each chromosome, a different kinship matrix was calculated where only the SNPs located on the other nine maize chromosomes were used to calculate the kinship matrix (Listgarten et al., 2012; Rincent et al., 2014).

The inspection of the inflation values for each model and the quantile-quantile (Q-Q) plots of the respective P-values, allowed defining the best statistical model to fit the phenotypic data. Models with inflation factor near 1 are better and for quantile-quantile (Q-Q) plots, it is expected that few P-values will deviate from their expected distribution. The observed P-values from marker-trait associations were used to draw Manhattan plots where the $-\log_{10}$ P values of each SNP were plotted against their chromosomal positions. A liberal threshold of $-\log_{10}$ (P-value) = 4 was set to identify significant marker-trait associations. The effect of the minor frequency SNP variant, reported in relation to the most frequent allele reference, was calculated.

2.8 Post-GWAS procedures

A local linkage disequilibrium (LD) study was performed to define the chromosomal regions to search for candidate genes for the traits under analysis.

This procedure was done in two steps: In Step 1, the average intra-chromosomal LD was estimated as the squared correlation coefficient r^2, after correcting for population structure using the principal component scores from Eigenanalysis, as implemented in Genstat software. For this calculation, the same subset of 1,821 SNPs previously used for the genetic structure analysis was employed. LD decay was visualized per chromosome by plotting r^2 against the physical mapping distance in Mb. A liberal threshold for LD decay ($r^2 > 0.1$) was used to estimate the average genetic distance for which markers were considered to be no longer correlated. In Step 2, a genomic window around each SNP location significantly associated with the traits analyzed was established by subtracting and adding the average genetic distance for LD decay ($r^2 > 0.1$), estimated in Step 1. All the SNP markers located within those windows were then used to estimate the local LD decay. At this point, a stricter threshold of $r^2 > 0.2$ was considered. The markers' positions flanking each local LD block were further used as queries positions on the maize genome browser, via MaizeGDB (https://www.maizegdb.org/gbrowse/), to retrieve the list of candidate genes mapped within those genomic regions.

The genome sequence of the maize inbred line B73 (*Zea mays* B73 RefGen_v3) was used as the reference genome for candidate gene analyses (Schnable et al., 2009). The functional annotation of the genes under the identified genomic regions was retrieved via Phytozome (Goodstein et al., 2011, Phytozome 11, version AGPv3 - Zea mays Ensembl-18) using the gene model identifier as the query. KEGG: Kyoto Encyclopedia of Genes and Genomes database (Kanehisa & Goto, 2000) was used to retrieve information on the pathways where the candidate genes could be involved.

3 Results

3.1 Maize flour compositional and pasting properties traits variation

As shown in Table 2, where the quality traits variance components and heritabilities are presented, the highest percentage of variance was typically due to differences between the inbred lines (σ^2g), except for mean particle size (SIZEL), setback from trough viscosity (SB1) and setback from peak viscosity (SB2), where the error variance component was higher. The G × E interaction variance component (σ^2g×y) was more evident for traits related to maize flour pasting properties (viscosity parameters). Nevertheless, and for all traits analyzed, the variance component associated to differences between inbred lines was far greater than the variance component attributed to the effect of G × E interaction term (σ^2g / σ^2g×y > 1).

Fiber content had the highest heritability value (h^2 = 65%, across growing seasons), and setback from peak viscosity (SB2) the lowest (h^2 = 31%, across growing seasons) (Table 2). Detailed information on the collection of maize inbred lines phenotypic values (range and mean ± standard deviation) for quality traits evaluated in two growing seasons (2011 and 2012) can be found in Table S4.

Considering the data obtained across the two growing seasons, fiber and protein content appeared strongly and positively correlated as well as peak viscosity (PV) with breakdown viscosity (BD_SqRt); or final viscosity (FV) with setback from trough viscosity (SB1), being all of those pairwise comparisons with phenotypic (Figure 1) and genetic correlations coefficients superior to 0.8 (r > 0.8; Table S2 and Table S3).

Table 2. Variance components and broad sense heritability values for quality traits in wholemeal flour of a collection of 132 maize inbred lines evaluated in two growing seasons (2011 and 2012).

Trait	Variance components [2]							h^2 heritability [3](%)		
	σ^2p	σ^2g	σ^2g×y	σ^2error	σ^2g (%)	σ^2g×y (%)	σ^2error (%)	2011	2012	Across growing seasons
Protein (PR), in %	1.253	0.733	0.074	0.447	58	6	36	49	65	58
Fiber (FI), in %	0.045	0.029	0.001	0.014	65	3	32	52	66	65
Fat (FT), in %	0.050	0.028	0.004	0.018	55	8	37	52	55	55
Starch from lyophilized flour (STL), in %	5.636	3.064	0.000	2.572	54	0	46	55	43	54
Mean particle size in lyophilized flour (SIZEL), in µm	615.987	238.055	46.222	331.710	39	8	54	32	64	39
Peak (maximum) viscosity (PV), in cP	1140444.924	633464.838	337305.824	169674.262	56	30	15	62	72	56
Trough (minimum) viscosity (TV), in cP	274105.231	118572.456	37737.140	117795.635	43	14	43	44	63	43
Final viscosity (FV), in cP	1636670.931	674118.279	354128.068	608424.584	41	22	37	41	66	41
Breakdown [1] (BD_SqRt), in cP	167.218	103.016	36.195	28.007	62	22	17	58	73	62
Setback from trough viscosity (SB1), in cP	953260.470	328820.308	157925.956	466514.206	34	17	49	26	66	34
Setback from peak viscosity (SB2), in cP	821769.624	250954.864	188952.150	381862.610	31	23	46	39	66	31

[1] Breakdown values were squared-root transformed | [2] Percentage of variance attributed to the individual terms of the statistical model: σ^2p corresponds to the phenotypic variance; σ^2g corresponds to the genotypic variance; σ^2g×y corresponds to the interaction between inbred lines and growing seasons variance; σ^2error (%) corresponds to the error variance; σ^2g (%) corresponds to the percentage of variance attributed to difference among inbred lines; σ^2g×y (%) corresponds to the percentage of variance attributed to the interaction between inbred lines and growing seasons; σ^2error (%) corresponds to the percentage of variance attributed to the block, row, column, and residual terms which altogether compose the error variance | [3] h^2=broad sense heritability obtained by fitting inbred lines as random terms in the statistical model.

A strong positive phenotypic correlation was also detected between lyophilized and non-lyophilized maize inbred lines samples in what respects starch content and mean particle size (r = 0.81 and r = 0.77, respectively, P < 0.001, Table S2).

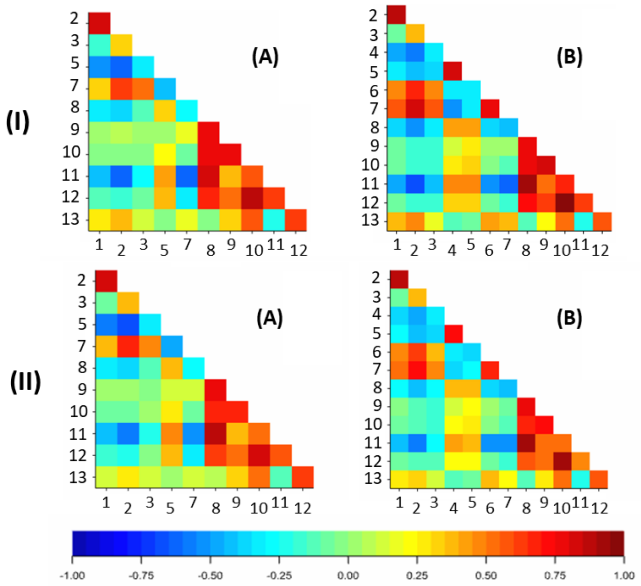

Figure 1. Heat maps illustrating the (I) phenotypic and (II) genetic correlations for compositional and pasting quality traits measured in wholemeal flour of 132 maize inbred lines grown during (A) 2011 growing season, and (B) 2012 growing season. Quality traits' key: 1 – Protein content, 2 – Fiber content, 3 – Fat content, 4 – Starch content in non-lyophilized flour, 5 – Starch content in lyophilized flour, 6 – Mean particle size in non-lyophilized flour, 7 – Mean particle size in lyophilized flour, 8 – Peak viscosity, 9 – Trough viscosity, 10 – Final viscosity, 11 – Breakdown viscosity (squared-root-transformed), 12 – Setback from trough viscosity, and 13 – Setback from peak viscosity.

The first two components of principal component analysis (explaining a total of 69.74% of the variability present in the dataset) depicted a high diversity among the inbred lines of the association panel for the quality traits analyzed (Figure 2). The maize inbred lines derived entirely from Portuguese traditional maize landraces were mainly located towards lower breakdown and peak viscosity values,

lower starch content and higher protein, fiber and mean particle size values.

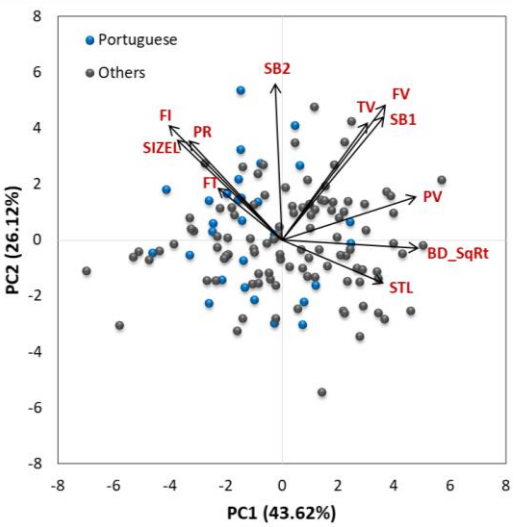

Figure 2. Principal component analysis (PCA) biplot based on BLUP values for 11 quality traits measured in 132 maize inbred lines. Circles colored in blue correspond to inbred lines selected entirely from Portuguese landraces. Quality traits' abbreviations: PR – percentage of protein; FI – percentage of fiber; FT – percentage of fat; STL – percentage of starch in lyophilized flour; SIZEL – mean particle size in lyophilized flour; PV – peak (maximum) viscosity; TV – trough (minimum) viscosity; FV – final viscosity; BD_SqRt – squared-root transformed values of the breakdown viscosity; SB1 – setback from trough viscosity; SB2 – setback from peak viscosity.

3.2 Genetic structure

From the performed principal components/Eigenanalysis (Figure 3), a wide dispersion of inbred lines was observed, with some separation of inbred lines according to kernel type (flint *vs.* dent) along PC1. The majority of the 29 lines selected directly from traditional Portuguese maize populations were clustered within the flint types (Figure S1).

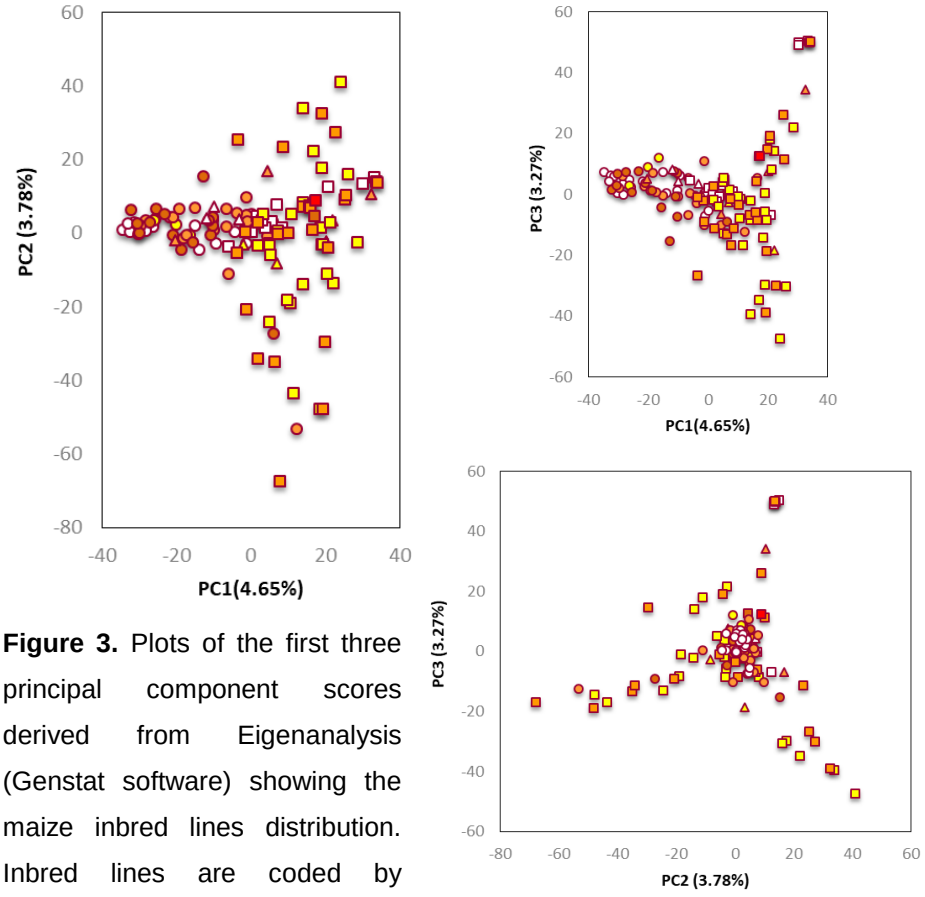

Figure 3. Plots of the first three principal component scores derived from Eigenanalysis (Genstat software) showing the maize inbred lines distribution. Inbred lines are coded by endosperm type: dent (squares), flint (circles), and intermediate (triangles) endosperms; and kernel color (white, yellow, yellow-orange, orange, and red). The variance explained by each principal component is given in the axis heading.

3.3 Genomic regions associated with quality traits

GWAS was performed using a mixed linear model (MLM) and either kinship relationship (K matrix) or population structure (Eigenanalysis) was taken into account to avoid spurious associations. After inspecting the observed inflation factors obtained for each tested model, the mixed linear model accounting for familial relatedness (K matrix) was selected as the best model (Table S5).

Therefore, the results reported below concern the results obtained using this model.

For all the studied major constituents of maize kernel (protein, fiber, fat, and starch content) and all the studied parameters affecting the maize flour technological performance (starch pasting properties and flour's mean particle size), significantly associated SNP markers were identified. In total, 72 unique SNPs were identified as being associated with the 11 quality traits analyzed across the two growing seasons (2011 and 2012) using a threshold $-\log_{10}$(P-value) = 4. The 72 SNPs corresponded to 57 genomic regions (LD $r^2 > 0.2$) (Figure 4, Table S6).

Considering the number of identified associated genomic regions across years per trait (Figure 4, also in the supplementary material Table S6 and Figure S2), breakdown viscosity (BD_SqRt) appeared as the trait with the bigger number of associated regions (nine regions, distributed among six different chromosomes), closely followed by protein content (PR), fiber content (FI) and mean particle size (SIZEL), all with eight associated genomic regions distributed respectively by three, four or five different chromosomes. Setback from trough viscosity (SB1) was the trait with the fewer detected associated regions (two regions, distributed between two different chromosomes). For all of the traits, and based on the rare allele contributions to the trait variation, SNPs associated with an increase as well as a decrease in the trait were detected (Table S7). The exception was trough viscosity (TV) and setback from peak viscosity (SB2) where the rare allele was always responsible for a decrease in the trait value (Table S7).

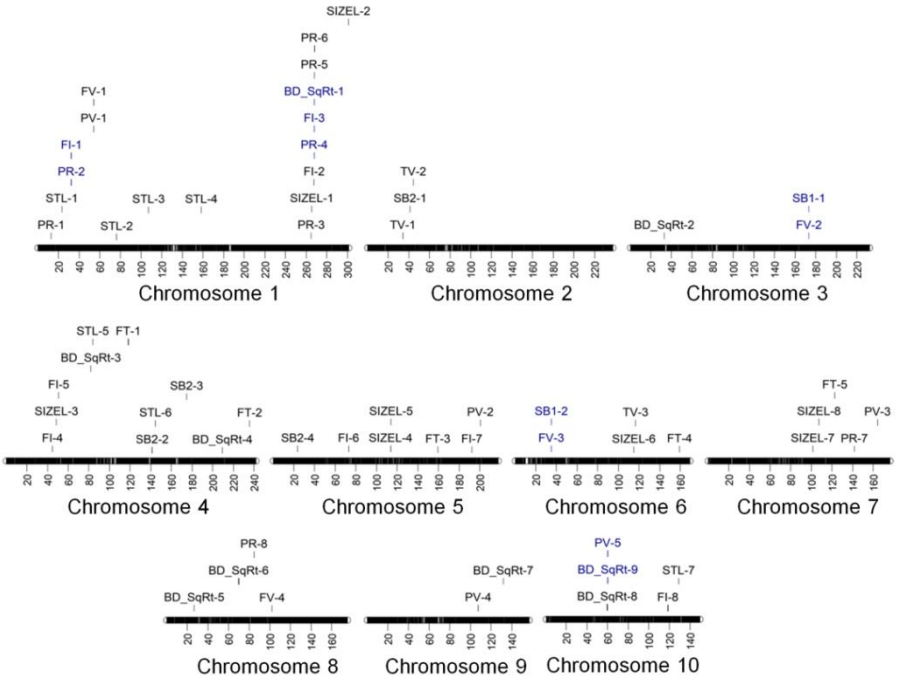

Figure 4. Schematic representation of the chromosomal regions identified by genome-wide association for the 11 quality traits using a collection of 132 maize inbred lines evaluated across two growing seasons. Horizontal bars represent each of the 10 maize chromosomes; for each chromosome, the SNP markers were sorted according to their positions, in megabase pairs. Each genomic region was termed accordingly to the trait, followed by a number identifying each individual region; vertical lines below correspond to the location of the genomic region associated with the trait variation. Co-localized regions associated with multiple traits are highlighted in blue. Traits abbreviations': PR – percentage of protein; FI – percentage of fiber; FT – percentage of fat; STL – percentage of starch in lyophilized flour; SIZEL – mean particle size in lyophilized flour ; PV – peak (maximum) viscosity; TV – trough (minimum) viscosity; FV – final viscosity; BD_SqRt – squared-root transformed values of the breakdown viscosity; SB1 – setback from trough viscosity; SB2 – setback from peak viscosity.

The significant SNP-traits associations only explained a small portion of the phenotypic variance observed for all traits, being that the range was smaller in the cases of fiber, protein, and starch content (percentage of phenotypic variance explained: for fiber, 6.46%-9.03%; for protein, 6.61%-9.88%; for starch, 6.79%-9.61%) (Table S7).

Some of these genomic regions were associated with multiple quality traits (Figure 4); many of those traits were highly correlated (Figure 1, Table S2 and Table S3). Protein (PR) and fiber content (FI) were simultaneously associated with two different genomic regions on chromosome 1 (Figure 5). One of those regions was located between 32,313 kilobase pairs (kb) and 32,548 kb and three significantly associated SNPs were identified. The other genomic region was located between 267,849 kb and 267,886 kb. This last region was also associated with breakdown viscosity (BD_SqRt) (Table S6).

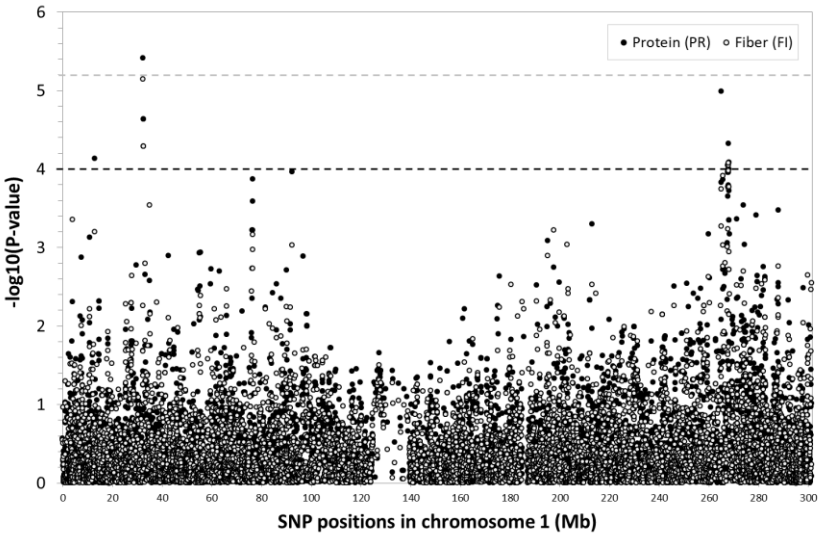

Figure 5. Chromosome 1 Manhattan plot with the genome-wide association results for protein and fiber content obtained using a collection of 132 maize inbred lines evaluated across two growing seasons. The *y*-axis shows the $-\log_{10}$ *P* values of 7,749 SNPs, and the *x*-axis shows their chromosomal

positions. Horizontal black and grey lines represent the liberal threshold of $P = 1\times10^{-4}$, and the Bonferroni-corrected threshold of $P = 6.45\times10^{-6}$, respectively.

Two other genomic regions were simultaneously associated with different traits more related to flour's pasting properties (traits from viscosity profiles). Namely, one genomic region associated with breakdown viscosity (BD_SqRt) and peak viscosity (PV) was identified on chromosome 10 (60,092 kb to 60,351 kb) (Figure 6). Two other regions were associated both to setback from trough viscosity (SB1) and final viscosity (FV) in chromosome 3 (173.419 kb to 173,420 kb) and chromosome 6 (34,978 kb to 35,091 kb) (Figure 7).

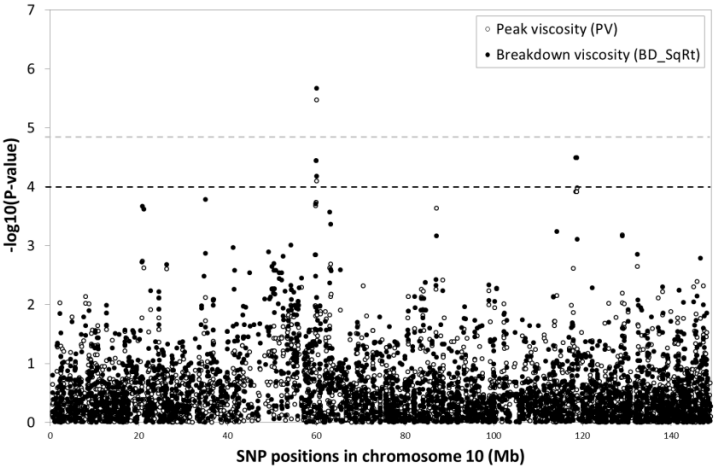

Figure 6. Chromosome 10 Manhattan plots with the genome-wide association results for peak viscosity (PV) and breakdown viscosity (BD_SqRt) obtained using a collection of 132 maize inbred lines evaluated across two growing seasons. The *y*-axis shows the $-\log_{10} P$ values of 3,477 SNPs, and the *x*-axis shows their chromosomal positions. Horizontal black and grey lines represent the liberal threshold of $P = 1\times10^{-4}$, and the Bonferroni-corrected threshold of $P = 1.44\times10^{-5}$, respectively.

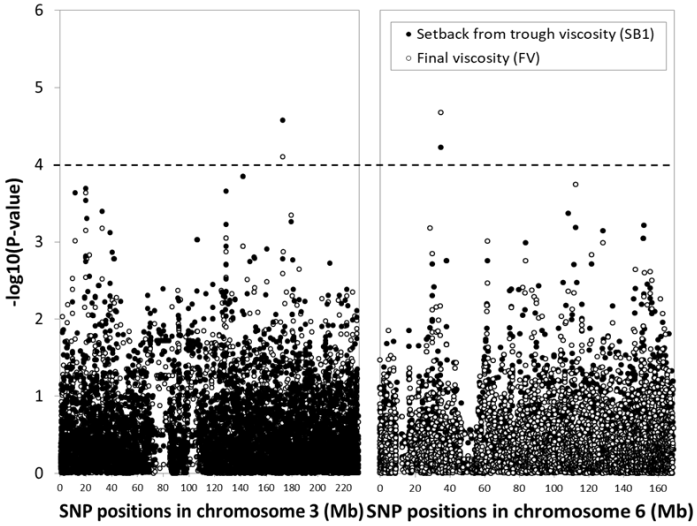

Figure 7. Chromosome 3 and chromosome 6 Manhattan plots with the genome-wide association results for setback from trough viscosity (SB1) and final viscosity (FV) obtained using a collection of 132 maize inbred lines evaluated across two growing seasons. The *y*-axis shows the $-\log_{10} P$ values of each SNP, and the *x*-axis shows their chromosomal positions. The horizontal black line represents the liberal threshold of $P = 1\times10^{-4}$.

Regions strongly associated to a single trait were also detected (Figure 4). This was the case of a genomic region on chromosome 5 strongly associated with the setback from peak viscosity (SB2) (located at 23,783 kb) (Table S6).

When inspecting the Manhattan plots per trait (Figure S2), it was also possible to identify regions where many different SNPs were associated consistently with the same trait although with a lower statistical level of significance. This was the case for fiber content (FI) with two associated genomic region, one on chromosome 1 (267,638 kb to 267,677 kb) and another on chromosome 10 (118.527 kb to 118,852 kb) respectively where four and three different associated SNP markers were located. Also on chromosome 1, two very close genomic regions (267,974 kb, and 268,031 kb to 268,218 kb) bared 3

different SNP markers associated with protein content (PR) (Table S6). For setback from peak viscosity (SB2), one of these regions was also detected in chromosome 4 (114,121 to 114,298 kb) where three different SNP markers were simultaneously associated with this trait variation (Table S6). Finally, and for mean particle size (SIZEL), two very close genomic regions, located on chromosome 5 (113,716 to 113,814 kb and 114,124 to 114,136 kb), bared 3 different SNPs associated with this trait variation (Table S6).

In summary, while some of the genomic regions were associated with several traits, the majority of the genomic regions were associated to a single trait.

3.4 Candidate genes identification

The average LD decay for the quality traits significantly associated genomic regions was 52.23 kb for LD $r^2 > 0.2$. This value extended to a maximum of 457 kb in a region of chromosome 10 spanning from 59,574 kb to 60,031 kb identified as being associated with breakdown viscosity trait (BD_SqRt) (Table S6). Using as reference the filtered gene set from the B73 RefGen_v3 assembly, a complete list of genes mapped within the significantly associated genomic regions identified in the GWAS for the 11 quality traits can be found in Table S8. A substantial proportion (66.67%) of the SNPs significantly associated with the quality traits were mapped within genes (48 out of 72 SNPs significantly associated with any trait, Table S8). And the degree of linkage disequilibrium around the genomic regions identified by GWAS allowed achieving a mapping resolution to the gene level for 40.35% of the cases (LD blocks where a single gene was identified, Table S8).

In the frame of this thesis, it was not possible to describe all candidate genes located within the associated genomic regions in

detail (Table S8). We here, therefore, restrict ourselves to describe those that were (1) located within regions where the strongest significant associations where detected, or (2) located within regions associated with multiple quality traits.

Genes within the regions with the strongest SNP-trait associations

The strongest SNP traits associations detected, corresponding to three different genomic regions, were located on chromosomes 1, 5, and 10 (SNPs highlighted in Table S6). One genomic region on chromosome 1 (32,314 kb to 32,548 kb) associated with protein content (PR), a second genomic region on chromosome 5 (23,783,411 bp) associated with setback from peak viscosity (SB2), and the last one on chromosome 10 (60,092 kb to 60,351 kb) associated with peak viscosity (PV) and breakdown viscosity (BD_SqRt).

In the genomic region identified on chromosome 1 (32,314 kb to 32,548 kb), three SNPs were significantly associated with protein (PR) content (rs131232105, rs131177502, and rs131232195). In this region, the strongest SNP associated with protein content, rs131232105 (–log10 (P-value) = 5.416), was located within the GRMZM2G099528 gene, coding for a B-cell receptor-associated 31-like protein. The other two significantly associated SNPs (rs131177502 and rs131232195, $-\log_{10}$ (P-value) = 4.636) were located within the GRMZM5G849530 gene, coding for a protein of unknown function. By considering the LD decay in that region, other associated genes could also be identified: GRMZM2G085427, coding for TSL-kinase interacting protein 1 (ZmMYBR59 transcription factor), GRMZM5G884325, coding for a small nuclear ribonucleoprotein Sm D3, involved in the spliceosome, GRMZM2G104255, coding for a member of the CRAMPED PROTEIN family (PTHR21677),

GRMZM5G868062, coding for a 60S ribosomal protein 15.5kD/SNU13, NHP2/L7A family (includes ribonuclease P subunit p38), involved in splicing. The genes mapped within the region spanning on chromosome 1 from 32,314 kb to 32,548 kb, besides being putatively involved in protein content, were also candidate genes for fiber content.

In the genomic region identified on chromosome 5, one significant SNP was strongly associated with setback from peak viscosity (SB2) (rs131504732, $-\log_{10}$ (P-value) = 5.846). This SNP was located within the GRMZM2G376743 gene, coding for a protein from the ARM repeat superfamily (PTHR33836:SF1). This SNP was not in LD ($r^2 > 0.2$) with its neighbor markers.

Finally, in the genomic region identified on chromosome 10, two SNPs were significantly associated with peak viscosity (PV) and breakdown viscosity (BD_SqRt) (rs128531960 and rs131765763). Of those SNPs, the strongest SNP associated with both traits was rs131765763 (–log10 (P-value) = 5.468, for peak viscosity, and (–log10 (P-value) = 5.671, for breakdown viscosity). The SNPs associated with those two traits were not mapped within any gene. Nevertheless, considering the LD decay around those SNPs, several genes were identified within the region: GRMZM2G079777, coding for a V-type proton ATPase subunit D protein, involved in the phagosome and in oxidative phosphorylation, GRMZM2G181192 (*glx1*), coding for glyoxylase1, involved in the pyruvate metabolism, GRMZM2G079925 and GRMZM2G005938, both coding for pentatricopeptide repeat-containing proteins.

Genes within regions associated with multiple quality traits

On chromosome 1, besides the region described in the previous section strongly associated with protein content, several candidate

genes were located on another genomic region identified being simultaneously associated with breakdown viscosity (BD_SqRt), fiber (FI) and protein (PR) content. This region spanned from 267,849 kb to 267,886 kb. One of the two significant SNPs for fiber content (rs131184056, $-\log_{10}$ (P-value) = 4.047) was mapped within the GRMZM2G127656 gene, which encodes a protein containing a zinc-finger domain of monoamine-oxidase A repressor R1. One of the two significant SNPs for protein content (rs128946745, $-\log_{10}$ (P-value) = 4.065) was mapped within the GRMZM2G022787 gene, which encodes for a pentatricopeptide repeat-containing protein. Considering the LD decay around the significant SNPs, the AC186684.4_FG001 gene, which encodes for a protein of unknown function, was also within that region.

Two regions were simultaneously associated with the final viscosity (FV), and setback from trough viscosity (SB1), the region of chromosome 3 spanning from 173,419 to 173,420 kb and the region on chromosome 6 spanning from 34,978 kb to 35,091 kb. In the first region (chromosome 3), the significant SNP (rs131180967, $-\log_{10}$ (P-value) = 4.571, for FV, and $-\log_{10}$ (P-value) = 4.103, for SB1) was mapped within the GRMZM2G452630 gene, coding for a serine hydroxyl-methyl-transferase related protein. In the second region (chromosome 6), the significant SNP (rs131176534, $-\log_{10}$ (P-value) = 4.675, for FV, and $-\log_{10}$ (P-value) = 4.224, for SB1) was mapped within the GRMZM2G045971 gene, coding for a preprotein translocase Sec Sec61-beta subunit protein. Other genes mapped within this last region, considering the LD decay around the significant SNPs, were GRMZM2G336583 gene, coding for a phragmoplastin interacting protein 1, GRMZM5G868296 gene, coding for a protein of unknown function, and GRMZM2G001205 gene, coding for a C2H2-type zinc finger protein.

4 Discussion

This work reports the identification of 57 genomic regions associated with the 11 different quality traits evaluated in wholemeal maize flour. This was achieved through the genome-wide association analysis that we undertook, using for the first time an original association panel containing inbred lines derived from traditional Portuguese maize populations. This study allowed to identify candidate genes for the majority of the quality associated genomic regions controlling for maize kernel main compositional and pasting quality variation. However also novel regions, with no clear candidates, were identified that were not previously acknowledged using other germplasm collections studies.

4.1 Genomic regions associated with flour composition and pasting properties

The number of regions identified for each quality trait varied from nine regions for breakdown viscosity (BD_SqRt), to two regions for setback from trough viscosity (SB1). Additionally, several regions controlling multiple traits were also identified, which was not surprising given the strong pairwise phenotypic and genotypic correlation detected between some of the traits evaluated (such as peak viscosity and breakdown viscosity, final viscosity and setback from tough viscosity or protein and fiber content). This detection of genomic regions associated with multiple traits variation could be due to pleiotropic effects, with a single gene affecting multiple traits. However, since several genes are mapped within some of those regions, as mentioned in Karn et al. (2017), fine mapping within these regions is still required in order to properly address if a pleiotropic gene is responsible for both traits variation or the traits' variation is

due to two closely linked genes, and investigate about the possibility of independent selection among the correlated traits.

In what concerns the 29 maize inbred lines derived directly from Portuguese populations, the multivariate analysis showed that the inbred lines derived from Portuguese maize populations were overall characterized by having lower breakdown viscosity, peak viscosity, and starch content, and higher protein, fiber and mean particle size values. Considering the effect of the most frequent allele of the strongest SNPs associated with those traits and/or the SNPs that explained the biggest proportion of genetic variance, we observed that indeed the frequency of the SNP variants in the Portuguese inbred lines were directed towards an increase in protein, fiber and mean particle size, and a decrease in starch, breakdown viscosity and peak viscosity. This can indicate a positive selection towards the presence of the favourable alleles for protein content (SNP rs131232105); for fiber content (SNP rs132587158) and mean particle size (SNP rs131635762), and alleles associated with a decrease in breakdown and peak viscosities values (SNP rs128531960), and decrease in starch content (SNP rs131186983) in the Portuguese maize germplasm. For example, the strongest associated SNP in chromosome 1 for protein content (rs131232105) (Table S7), the variant allele had an effect on the reduction of protein content (−0.56%) in comparison to the most frequent allele. The same SNP was also associated with fiber content, and also, in this case, the variant allele had an effect on the reduction of fiber content (−0.10%) in comparison to the most frequent allele. We observe in the 29 inbred lines derived from entirely from Portuguese maize populations that the unfavorable allele was only present in ~10% of the Portuguese lines.

4.2 Candidate genes identification

Several of the SNP-trait associations detected in the present study were located within or near *a priori* candidate gene, which strengthened and served as a proof-of-concept for the usefulness of the used association panel, thought the statistical power to detect the significant associations was clearly constrained most likely by the size of the association panel and by the fast LD decay rate observed in the regions associated with the traits analyzed.

Some of the genomic regions identified in this work harbored potential candidate genes for which we had no previous information on their involvement with the quality traits analyzed. This was the case for one of the genomic regions on chromosome 1 strongly associated with protein content and also associated with fiber content. These “novel” regions with unforeseen candidate genes, not previously described as associated to the studied traits, may be due to the use of different association panels harboring different genetic variability, or simply be due to the rapid rate of LD decay observed in the present panel that hampered the identification of the obvious candidate.

For one of the flour pasting properties studied in this work, associated with breakdown viscosity, a promising *a priori* candidate gene − *dull endosperm 1* gene (GRMZM2G141399, *du1*) − was located near two identified associated genomic regions on chromosome 10 at a distance of ~46 kb and ~564 kb downstream of the confidence intervals considered (59,574-60,031 kb and 60,092-60,351 kb, respectively). The *dull endosperm 1* gene encodes a starch synthase and is a determinant of the structure of endosperm starch in maize (Gao et al., 1998; Wu et al., 2015).

As mentioned by Bian et al. (2014), GWAS associations have higher resolution, but often lower power due to stringent thresholds designed to minimize false positive associations, and leading to variability of detection across studies. For instance, in the current work the allelic variation for SNP ID rs128946933 (chromosome 1 at 267,974,184 bp was almost below the threshold considered to for a significant SNP-trait association ($-\log_{10}$ (P-value) = 4.002). This SNP was associated with protein content in maize flour and was located within the GRMZM2G066749 gene. Recently, Chen et al. (2017) demonstrated that this particular GRMZM2G066749 gene is the causative gene for *dek35* mutants. The mature *dek35* seeds exhibited a significant decrease in seed dry weight and zein protein content (Chen et al., 2017).

Several other *a priori* candidate genes previously identified as associated with maize kernel composition traits and starch pasting properties (e.g., Cook et al., 2012; Wilson et al., 2004; Xu et al., 2014) were not detected in the present study. Examples are the *brittle endosperm2* (*bt2*), *shrunken1* (*sh1*), and s*hrunken2* (*sh2*) known for significant association with kernel composition traits, as well as the *amylose extender1* (*ae1*) and *sh2* known for significant association with starch pasting properties (Wilson et al., 2004). Another example of a potential candidate gene for starch content would be the *brittle endosperm1* gene (GRMZM2G144081, *bt1*), coding a protein Brittle1 (Bt1) protein, involved in ADP-Glc transport into endosperm plastids and playing a role in starch biosynthesis (Xu et al., 2014). Associated with oil content in maize kernels is the acyl-CoA:diacylglycerol acyltransferase gene (GRMZM2G169089, *DGAT1-2*), (Cook et al, 2012; Zheng et al., 2008). The genotyping platform used on the current work screened several SNPs located within all the aforementioned candidate genes. Nevertheless, no association was

detected between those SNPs and the level maize kernel composition traits and starch pasting properties on the present association panel. As pointed out by Cook et al. (2012) several factors could be responsible for differences in position and quantity of QTLs detected between studies, including variation in allelic frequency, mapping resolution influenced by the magnitude of linkage disequilibrium in a population, marker density, environmental effects, and QTL analysis methods (Cook et al., 2012).The relatively small size of the used association panel might have constrained the statistical power to detect significant marker-trait associations in the present study.

An ideal association panel should harbor as much genetic diversity as possible, which would be used to resolve complex trait variation to a single gene or nucleotide (Yang et al., 2010). However, the genetic diversity should also be balanced with the genetic homogeneity of phenotypic traits, to ensure equal adaptation of all lines in multiple environments for phenotypic data collection (Yang et al., 2010). The rapid rate of LD decay observed in the present study in the SNPs associated with the quality traits evaluated suggests that a higher marker density would have been beneficial in the detection of other regions putatively linked to maize flour's quality.

Findings from the GWAS provide valuable genetic information into trait architecture or candidate loci for subsequent validation (Korte & Farlow, 2013). Preliminary GWAS analysis should be complemented by statistical procedures to help prioritize GWAS results (Cantor et al., 2010), as well as other follow-up analyses of GWAS loci and additional experiments, may be required to pinpoint the causal genes (Huang & Han, 2014). The SNPs strongly associated with the traits analyzed, and/or the SNPs which allelic variant was found to contribute to larger phenotypic effects, should be

prioritized as candidate genomic regions for marker development to support selection activities especially for the quality-related traits difficult to measure/assess. Nevertheless, prior to that, those associations need to be validated. Future work will concentrate on the validation of the results retrieved in this work by sequencing those regions on contrasting maize populations for the given trait. Since the actual materials used for the manufacturing of maize-based bread are the maize populations, these are the ideal independent materials to proceed with the missing validation.

Acknowledgments

MLA performed the DNA isolation, the genotypic and phenotypic data analysis, the association mapping analysis and follow-up analysis, and drafted the manuscript. BC performed the quality data acquisition. DG, CB, and MP participated in the establishment and maintenance of the field trials and performed the controlled pollinations. PMM coordinated the field trials and participated in the manuscript revision. CB coordinated and participated in the acquisition of quality data, and participated in the manuscript revision. MM and FvE participated in the molecular and phenotypic data processing, in the genome-wide association analysis, and participated in the manuscript revision. MCVP designed and coordinated the study and participated in the drafting and revising of the manuscript.

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) and Horizon 2020 Framework Programme under grant agreements No 245058 and No 633571, respectively, and from the Fundação para a Ciência e a Tecnologia, through the research unit GREEN-IT (UID/Multi/04551/2013), and through the research project under the

grant agreement no. PTDC/AGR-ALI/099285/2008. MLA and MCVP were supported by Fundação para a Ciência e a Tecnologia (SFRH/BD/77723/2011; FCT Investigator 2015 contract - IF/01337/2014, respectively). The authors would also like to acknowledge the Portuguese Bank of Plant Germplasm for providing the original seeds of the maize inbred lines.

References

Almeida-Dominguez, H. D., Suhendro, E. L., & Rooney, L. W. (1997). Factors affecting rapid visco analyser curves for the determination of maize kernel hardness. *Journal of Cereal Science*, 25(1), 93-102.

American Association of Cereal Chemists – AACC. (1999). Method 55-40.01:1999: approved methods of analysis: particle size of wheat flour by laser instrument (11th ed.). St. Paul.

Bian, Y., Yang, Q., Balint-Kurti, P. J., Wisser, R. J., & Holland, J. B. (2014). Limits on the reproducibility of marker associations with southern leaf blight resistance in the maize nested association mapping population. *BMC Genomics*, 15(1), 1068.

Brites, C., Trigo, M. J., Santos, C., Collar, C., & Rosell, C. M. (2010). Maize-based gluten-free bread: influence of processing parameters on sensory and instrumental quality. *Food and Bioprocess Technology*, 3(5), 707-715.

Cantor, R. M., Lange, K., & Sinsheimer, J. S. (2010). Prioritizing GWAS results: a review of statistical methods and recommendations for their application. *The American Journal of Human Genetics*, 86(1), 6-22.

Carbas, B., Vaz Patto, M. C., Bronze, M. R., Bento da Silva, A., Trigo, M. J., & Brites, C. (2016). Maize flour parameters that are related to the consumer perceived quality of 'broa' specialty bread. *Food Science and Technology*, 36, 259–267.

Chen, M., Rao, R. S. P., Zhang, Y., Zhong, C., & Thelen, J. J. (2016). Metabolite variation in hybrid corn grain from a large-scale multisite study. *The Crop Journal*, 4(3), 177-187.

Chen, X., Feng, F., Qi, W., Xu, L., Yao, D., Wang, Q., et al. (2017). Dek35 encodes a PPR protein that affects cis-splicing of mitochondrial nad4 intron 1 and seed development in maize. *Molecular Plant*, 10(3), 427-441.

Cook, J. P., McMullen, M. D., Holland, J. B., Tian, F., Bradbury, P., Ross-Ibarra, J., et al. (2012). Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. *Plant Physiology*, 158(2), 824-834.

Food and Agriculture Organization of the United Nations (FAO). (2012). FAOSTAT. Rome, Italy: FAO.

Ganal, M. W., Durstewitz, G., Polley, A., Bérard, A., Buckler, E. S., Charcosset, A., et al. (2011). A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PloS one*, 6(12), e28334.

Gao, M., Wanat, J., Stinard, P. S., James, M. G., & Myers, A. M. (1998). Characterization of *dull1*, a maize gene coding for a novel starch synthase. *The Plant Cell*, 10(3), 399-412.

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2011). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*, 40(D1), D1178-D1186.

Gwirtz, J. A., & Garcia-Casal, M. N. (2014). Processing maize flour and corn meal food products. *Annals of the New York Academy of Sciences*, 1312(1), 66-75.

Hartings, H., Fracassetti, M., & Motto, M. (2012). Genetic enhancement of grain quality-related traits in maize. In *Transgenic Plants-Advances and Limitations*, Y. Çiftçi (Ed.), (pp. 191–218), Rijeka,Croatia: InTech.

Huang, X., & Han, B. (2014). Natural variations and genome-wide association studies in crop plants. *Annual Review of Plant Biology*, 65, 531-551.

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1), 27-30.

Karn, A., Gillman, J. D., & Flint-Garcia, S. A. (2017). Genetic analysis of teosinte alleles for kernel composition traits in maize. *G3: Genes, Genomes, Genetics*, 7(4), 1157-1164.

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: a review. *Plant Methods*, 9(1), 29.

Lawrence, C. J., Harper, L. C., Schaeffer, M. L., Sen, T. Z., Seigfried, T. E., & Campbell, D. A. (2008). MaizeGDB: The Maize Model Organism Database for Basic, Translational, and Applied Research. *International Journal of Plant Genomics*, 2008:496957.

Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., et al. (2013). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nature genetics*, 45(1), 43-50.

Listgarten, J., Lippert, C., Kadie, C. M., Davidson, R. I., Eskin, E., & Heckerman, D. (2012). Improved linear mixed models for genome-wide association studies. *Nature Methods*, 9(6), 525-526.

Malosetti, M., van der Linden, C. G., Vosman, B., & van Eeuwijk, F. A. (2007). A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. *Genetics*, 175(2), 879-889.

Mendes-Moreira, P., Alves, M. L., Satovic, Z., dos Santos, J. P., Santos, J. N., Souza, J. C., et al. (2015). Genetic architecture of ear fasciation in maize (*Zea mays*) under QTL scrutiny. *PloS one*, 10(4), e0124543.

Murray, D., Payne, R, & Zhang, Z. (2014) Breeding View, a Visual Tool for Running Analytical Pipelines: User Guide. VSN International Ltd.

Paraginski, R. T., Vanier, N. L., Berrios, J. D. J., De Oliveira, M., & Elias, M. C. (2014). Physicochemical and pasting properties of maize as affected by storage temperature. *Journal of Stored Products Research*, 59, 209-214.

Payne R., Murray D., and Harding S. (2015). GenStat for Windows (18th Edition) Introduction. VSN International: Hemel Hempstead, UK.

Ranum, P., Peña-Rosas, J. P., & Garcia-Casal, M. N. (2014). Global maize production, utilization, and consumption. *Annals of the New York Academy of Sciences*, 1312(1), 105-112.

Rincent, R., Moreau, L., Monod, H., Kuhn, E., Melchinger, A. E., Malvar, R. A., et al. (2014). Recovering power in association mapping panels with variable levels of linkage disequilibrium. *Genetics*, 197(1), 375-387.

Saghai-Maroof, M.A., Soliman, K.M., Jorgensen, R.A., & Allard, R.W. (1984). Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics.

Proceedings of the National Academy of Sciences of the United States of America, 81(24), 8014-8018.

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science*, 326(5956), 1112-1115.

Tenaillon, M.I., & Charcosset, A. (2011). A European perspective on maize history. *Comptes Rendus Biologies*, 334(3), 221-228.

The IBP Breeding Management System Version 3.0.9 (December 2015). The Integrated Breeding Platform. Available at <https://www.integratedbreeding.net/breeding-management-system>

Vaz Patto, M. C., Moreira, P. M., Carvalho, V., & Pego, S. (2007). Collecting maize (*Zea mays* L. convar. *mays*) with potential technological ability for bread making in Portugal. *Genetic Resources and Crop Evolution*, 54(7), 1555-1563.

Vaz Patto, M. C., Satovic, Z., Pego, S., & Fevereiro, P. (2004). Assessing the genetic diversity of Portuguese maize germplasm using microsatellite markers. *Euphytica*, 137(1), 63-72.

Wang, T., Wang, M., Hu, S., Xiao, Y., Tong, H., Pan, Q., et al. (2015). Genetic basis of maize kernel starch content revealed by high-density single nucleotide polymorphism markers in a recombinant inbred line population. *BMC Plant Biology*, 15(1), 288.

Wang, S., Zheng, M., Yu, J., Wang, S., & Copeland, L. (2017). Insights into the Formation and Structures of Starch–Protein–Lipid Complexes. *Journal of Agricultural and Food Chemistry*, 65(9), 1960-1966.

Wilson, L. M., Whitt, S. R., Ibáñez, A. M., Rocheford, T. R., Goodman, M. M., & Buckler, E. S. (2004). Dissection of maize kernel composition and starch production by candidate gene association. *The Plant Cell*, 16(10), 2719-2733.

Wu, J. D., Jiang, C. P., Zhu, H. S., Jiang, H. Y., Cheng, B. J., & Zhu, S. W. (2015). Cloning and functional analysis of the promoter of a maize starch synthase III gene (*ZmDULL1*). *Genetics and Molecular Research*, 14, 5468-5479.

Xu, S., Yang, Z., Zhang, E., Jiang, Y., Pan, L., Chen, Q., et al. (2014). Nucleotide diversity of maize ZmBT1 gene and association with starch physicochemical properties. *PloS one*, 9(8), e103627.

Yang, X., Yan, J., Shah, T., Warburton, M. L., Li, Q., Li, L., et al. (2010). Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. *Theoretical and Applied Genetics*, 121(3), 417-431.

Zhang, G., & Hamaker, B. R. (2003). A three component interaction among starch, protein, and free fatty acids revealed by pasting profiles. *Journal of Agricultural and Food Chemistry*, 51(9), 2797-2800.

Zheng, P., Allen, W. B., Roesler, K., Williams, M. E., Zhang, S., Li, J., et al. (2008). A phenylalanine in DGAT is a key determinant of oil content and composition in maize. *Nature Genetics*, 40(3), 367-372.

Supplementary material

Tables

Table S1. Maize inbred lines with available quality data, known pedigree, kernel color, and endosperm type. – Table available online through the link <https://figshare.com/s/34958547ba5cd15705bb>

Table S2. Pearson correlation coefficients among quality traits measured in wholemeal flour of 132 maize inbred lines. The phenotypic correlations were calculated independently for each growing season evaluated (2011 and 2012). The values above the diagonal correspond to the phenotypic correlations among quality traits measured in the first growing season (2011); values below the diagonal correspond to the phenotypic correlations among quality traits measured in the second growing season (2012). In bold are highlighted the strong phenotypic correlations (|r| > 0.8).

		PR	FI	FT	ST	STL	SIZE	SIZEL	PV	TV	FV	BD_SqRt	SB1	SB2
PR	r	-	**0.82**	-0.17	NA	-0.51	NA	0.32	-0.30	0.02	-0.04	-0.43	-0.11	0.29
	P-value	-	***	ns	NA	***	NA	**	**	ns	ns	***	ns	**
FI	r	**0.87**	-	0.33	NA	-0.62	NA	0.65	-0.37	0.10	-0.04	-0.61	-0.17	0.39
	P-value	***	-	***	NA	***	NA	***	***	ns	ns	***	ns	***
FT	r	-0.07	0.39	-	NA	-0.32	NA	0.52	-0.15	0.04	-0.04	-0.27	-0.09	0.12
	P-value	ns	***	-	NA	**	NA	***	ns	ns	ns	**	ns	ns
ST	r	-0.46	-0.55	-0.33	-	NA	NA	NA	NA	NA	NA	NA	NA	NA
	P-value	***	***	**	-	NA	NA	NA	NA	NA	NA	NA	NA	NA
STL	r	-0.31	-0.44	-0.35	**0.81**	-	NA	-0.41	0.31	0.03	0.22	0.45	0.35	-0.03
	P-value	**	***	**	***	-	NA	***	**	ns	*	***	***	ns
SIZE	r	0.46	0.68	0.47	-0.32	-0.32	-	NA	NA	NA	NA	NA	NA	NA
	P-value	***	***	***	**	**	-	NA	NA	NA	NA	NA	NA	NA
SIZEL	r	0.59	**0.80**	0.56	-0.53	-0.34	**0.77**	-	-0.29	0.20	-0.10	-0.60	-0.31	0.20
	P-value	***	***	***	***	**	***	-	***	ns	ns	***	**	ns
PV	r	-0.37	-0.52	-0.33	0.42	0.45	-0.36	-0.42	-	**0.77**	**0.77**	**0.85**	0.68	-0.01
	P-value	**	***	**	***	***	**	***	-	***	***	***	***	ns
TV	r	-0.07	-0.15	-0.17	0.11	0.29	0.01	0.01	**0.76**	-	**0.76**	0.37	0.54	0.31
	P-value	ns	ns	ns	ns	*	ns	ns	***	-	***	***	***	**
FV	r	-0.07	-0.17	-0.19	0.27	0.34	-0.05	-0.13	**0.78**	**0.81**	-	0.55	**0.89**	0.56
	P-value	ns	ns	ns	*	**	ns	ns	***	***	-	***	***	***

Continuation of Table S2

		PR	FI	FT	ST	STL	SIZE	SIZEL	PV	TV	FV	BD_SqRt	SB1	SB2
BD_SqRt	r	-0.48	-0.67	-0.42	0.48	0.45	-0.54	-0.61	**0.93**	0.53	0.66	-	0.61	-0.20
	P-value	***	***	***	***	***	***	***	***	***	***	-	***	ns
SB1	r	-0.06	-0.15	-0.17	0.32	0.32	-0.07	-0.19	0.71	0.61	**0.96**	0.64	-	0.63
	P-value	ns	ns	ns	**	**	ns	ns	***	***	***	***	-	***
SB2	r	0.40	0.45	0.16	-0.14	-0.06	0.45	0.37	-0.11	0.25	0.50	-0.23	0.56	-
	P-value	***	***	ns	ns	ns	***	**	ns	*	***	ns	***	-

*r corresponds to the Person correlation coefficient; P-value corresponds to the significance level of correlations indicated as: ns - non-significant; * - significant at P < 0.05; ** - significant at P < 0.01; *** - significant at P < 0.001 | NA corresponds to data not available. | Quality traits' abbreviations: PR – percentage of protein; FI – percentage of fiber; FT – percentage of fat; ST – percentage of starch in non-lyophilized flour; STL – percentage of starch in lyophilized flour; SIZE – mean particle size in non-lyophilized flour; SIZEL – mean particle size in lyophilized flour; PV – peak (maximum) viscosity; TV – trough (minimum) viscosity; FV – final viscosity; BD_SqRt – squared-root transformed values of the breakdown viscosity; SB1 – setback from trough viscosity; SB2 – setback from peak viscosity*

Table S3. Estimated genetic correlations among quality traits measured in wholemeal flour of a collection of 132 maize inbred lines evaluated during two growing seasons. The genetic correlations were calculated independently for each growing season (2011 and 2012). Values above the diagonal correspond to the genetic correlations among quality traits measured in the first growing season (2011); values below the diagonal correspond to the genetic correlations among quality traits measured in the second growing season (2012). In bold are highlighted the strong genetic correlations ($|r| > 0.8$).

	PR	FI	FT	STL	SIZEL	PV	TV	FV	BD_SqRt	SB1	SB2
PR	-	**0.802**	–0.075	–0.558	0.400	–0.281	–0.007	–0.051	–0.435	–0.063	0.264
FI	**0.855**	-	0.356	–0.664	0.669	–0.416	–0.110	–0.129	–0.571	–0.118	0.307
FT	–0.051	0.368	-	–0.335	0.491	–0.287	–0.173	–0.168	–0.286	–0.137	0.125
STL	–0.298	–0.438	–0.367	-	–0.487	0.375	0.229	0.278	0.383	0.244	–0.074
SIZEL	0.537	0.717	0.456	–0.297	-	–0.402	–0.060	–0.136	–0.545	–0.162	0.247
PV	–0.307	–0.385	–0.142	0.362	–0.272	-	**0.751**	0.690	**0.924**	0.616	0.071
TV	0.008	0.033	–0.029	0.113	0.122	**0.759**	-	0.665	0.531	0.517	0.320
FV	–0.086	–0.086	0.036	0.258	–0.096	0.669	**0.747**	-	0.527	**0.808**	0.507
BD_SqRt	–0.415	–0.576	–0.246	0.454	–0.508	**0.852**	0.395	0.504	-	0.554	–0.139
SB1	–0.187	–0.202	–0.040	0.351	–0.310	0.562	0.535	**0.950**	0.473	-	0.593
SB2	0.148	0.276	0.128	0.011	0.132	–0.101	0.266	0.524	–0.229	0.616	-

Quality traits' abbreviations: PR –percentage of protein; FI – percentage of fiber; FT – percentage of fat; STL – percentage of starch in lyophilized flour; SIZEL – mean particle size in lyophilized flour; PV – peak (maximum) viscosity; TV – trough (minimum) viscosity; FV – final viscosity; BD_SqRt – squared-root transformed values of the breakdown viscosity; SB1 – setback from trough viscosity; SB2 – setback from peak viscosity

Table S4. Phenotypic values (range and mean ± standard deviation) for quality traits in wholemeal flour of a collection of 132 maize inbred lines evaluated in two growing seasons (2011 and 2012).

Trait	Range (minimum - maximum)		Mean ± standard deviation		
	2011	2012	2011	2012	Across growing seasons
Protein (PR), in %	9.72 – 15.76	9.12 – 15.04	12.73 ± 1.09	12.07 ± 1.13	12.45 ± 1.15
Fiber (FI), in %	1.79 – 2.92	1.70 – 2.80	2.32 ± 0.21	2.29 ± 0.22	2.31 ± 0.21
Fat (FT), in %	4.02 – 5.19	4.30 – 5.39	4.61 ± 0.21	4.78 ± 0.22	4.69 ± 0.23
Starch from lyophilized flour (STL), in %	64.83 – 76.04	61.47 – 73.43	69.96 ± 2.17	67.57 ± 2.59	68.89 ± 2.65
Mean particle size in lyophilized flour (SIZEL), in µm	82.72 – 210.71	108.47 – 212.11	144.78 ± 26.17	164.22 ± 22.61	153.30 ± 26.47
Peak (maximum) viscosity (PV), in cP	836.50 – 5,303.00	551.00 – 5,926.00	3,104.81 ± 979.68	3,064.41 ± 1151.73	3,086.00 ± 1,063.00
Trough (minimum) viscosity (TV), in cP	637.00 – 3,424.00	830.00 – 3,494.00	1,980.00 ± 543.88	2,094.28 ± 504.38	2,031.00 ± 529.10
Final viscosity (FV), in cP	3,006.00 – 9,042.00	3,055.00 – 9,759.00	6,122.70 ± 1,190.71	6,285.31 ± 1,362.07	6,195.00 ± 1,270.00
Breakdown [1] (BD_SqRt), in cP	3.32 – 56.60	1.00 – 56.08	31.86 ± 11.18	27.74 ± 14.67	30.05 ± 12.98
Setback from trough viscosity (SB1), in cP	1,573.00 – 6,546.00	2,006.00 – 6,708.00	4,039.09 ± 932.82	4,162.82 ± 1,017.19	4,093.00 ± 971.00
Setback from peak viscosity (SB2), in cP	538.00 – 4,836.00	938.00 – 5,257.00	2,881.25 ± 894.89	3,096.36 ± 906.60	2,969.00 ± 904.60

[1] *Breakdown values were squared-root transformed*

Table S5. Observed inflation factors for the models tested in genome-wide association (GWAS) analysis. Inflation factor for the adaptive kinship model corresponds to the average value across chromosomes.

#	Trait	Naive	Eigen	Adaptive Kinship [1]
1	Protein content (PR)	1.409	1.133	1.062
2	Fiber content (FI)	1.730	1.080	1.079
3	Fat content (FT)	1.578	1.055	1.033
4	Starch content (STL)	1.658	1.125	1.091
5	Mean particle size (SIZEL)	1.647	1.105	1.108
6	Peak viscosity (PV)	1.690	1.077	1.048
7	Trought viscosity (TV)	1.270	1.064	1.019
8	Final viscosity (FV)	1.246	1.051	1.036
9	Breakdown viscosity (BD_SqRt)	1.782	1.098	1.090
10	Setback from trough viscosity (SB1)	1.138	1.038	1.038
11	Setback from peak viscosity (SB2)	1.149	1.053	1.036

[1] *Calculated according to Listgarten et al., 2012; Rincent et al., 2014*

Table S6. Significant SNP-trait associations using -log10 (P-value) = 4, as the threshold from a genome-wide association study for 11 quality traits using a collection of 132 maize inbred lines evaluated across two growing seasons. – Table available online through the link <https://figshare.com/s/d367b1be4d441c879144>

Table S7. Percentage of associated SNP variants with an effect of decrease and increase of the traits value, and maximum and minimum phenotypic variance explained by the SNPs associated with 11 quality traits.

%SNPs DECREASE – Percentage of associated SNPs for which the effect of the rare allele resulted in a decrease of the trait value when compared to the most frequent allele; %SNPs INCREASE - Percentage of associated SNPs for which the effect of the rare allele results in an increase of the trait value when compared to the most frequent allele; V_{MIN} – smallest proportion of variance explained by a significant SNP; V_{MAX} – largest proportion of variance explained by a significant SNP; SNP_{EFFECT} - Effect of variant allele explaining the largest proportion of the trait variance.

Trait	% SNPs DECREASE	%SNPs INCREASE	V_{MIN} (SNP ID; location)	V_{MAX} (SNP ID; location)	SNP_{EFFECT}
Protein (PR), in %	91.67	8.33	0.066 (rs131210797; chr1: 12,944,506 bp)	0.099 (rs131232105; chr1: 32,313,522 bp)	(−)0.56%
Fiber (FI), in %	93.75	6.25	0.065 (rs131178018, rs131180681, rs131180682, rs131180679; chr1: 267,638,749-267,677,992 bp)	0.090 (rs132587158; chr10: 118,706,425 bp)	(−)0.09%
Fat (FT), in %	40.00	60.00	0.079 (rs131178929; chr4: 235,742,784 bp)	0.102 (rs132129576; chr4: 118,410,565 bp)	(+)0.12%
Starch (STL), in %	57.14	42.86	0.068 (rs128781599; chr1: 158,498,846 bp)	0.096 (rs131186983; chr1: 23,242,656 bp)	(−)1.02%
Mean particle size (SIZEL), in µm	33.33	66.67	0.105 (rs131176855; chr6: 114,921,724 bp)	0.148 (rs131635762; chr7: 101,287,298 bp)	(−)18.13 µm
Peak viscosity (PV), in cP	33.33	66.67	0.080 (rs128531960; chr10: 60,163,242 bp)	0.115 (rs131765763; chr10: 60,163,343 bp)	(−)383.98 cP

Continuation Table S7

Trait	% SNPs DECREASE	%SNPs INCREASE	V_{MIN} (SNP ID; location)	V_{MAX} (SNP ID; location)	SNP_{EFFECT}
Trough viscosity (TV), in cP	100.00	0.00	0.094 (rs130376196; chr6: 116,511,199 bp)	0.118 (rs129057461; chr2: 44,381,952 bp)	(−)184.4 cP
Final viscosity (FV), in cP	50.00	50.00	0.096 (rs130828069; chr8: 102,473,765 bp)	0.111 (rs131180967; chr3: 173,420,781 bp)	(−)422.1 cP
Breakdown viscosity (BD_SqRt), in cP	45.45	54.55	0.073 (rs131664622; chr8: 26,757,485 bp)	0.112 (rs131765763; chr10: 60,163,343 bp)	(−)4.82 cP
Setback from trough viscosity (SB1), in cP	50.00	50.00	0.111 (rs131176534; chr6: 35,091,373 bp)	0.113 (rs131180967; chr3: 173,420,781 bp)	(−)298.93 cP
Setback from peak viscosity (SB2), in cP	100.00	0.00	0.133 (rs131182722, rs131182721, rs132138984; chr4: 141,121,126-141,298,543 bp)	0.156 (rs132163268; chr4: 174,747,548 bp)	(−)503.09 cP

Table S8. Candidate genes mapped within the genomic regions associated with 11 quality traits. – Table available online through the link < https://figshare.com/s/e163a933dc8cfc1e08aa>

Figures

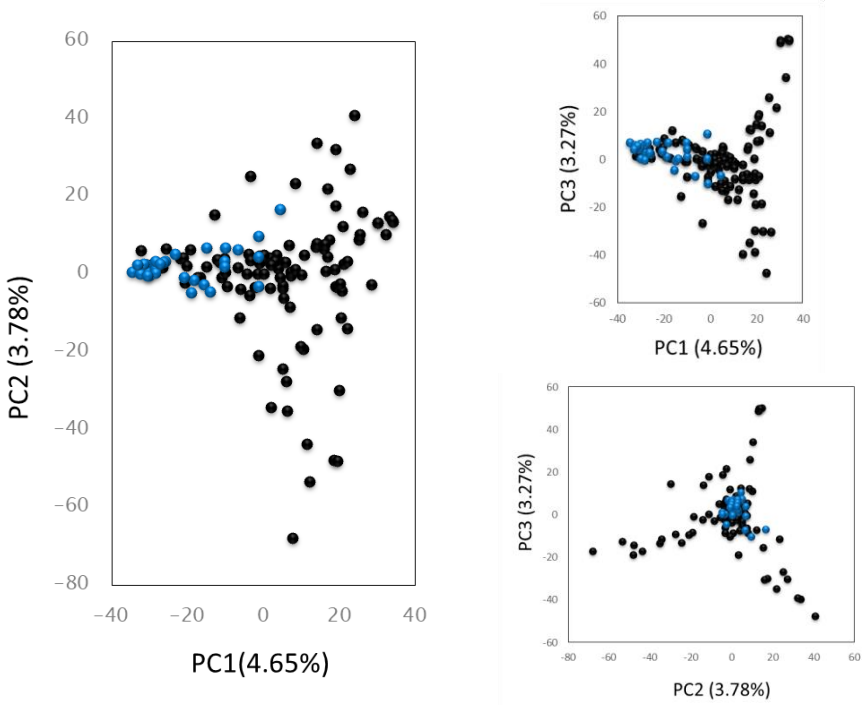

Figure S1. Plots of the first three principal component scores derived from Eigenanalysis (Genstat software) showing the maize inbred lines distribution. Inbred lines selected directly from traditional Portuguese maize populations are depicted in blue. The variance explained by each principal component is given in the axis heading.

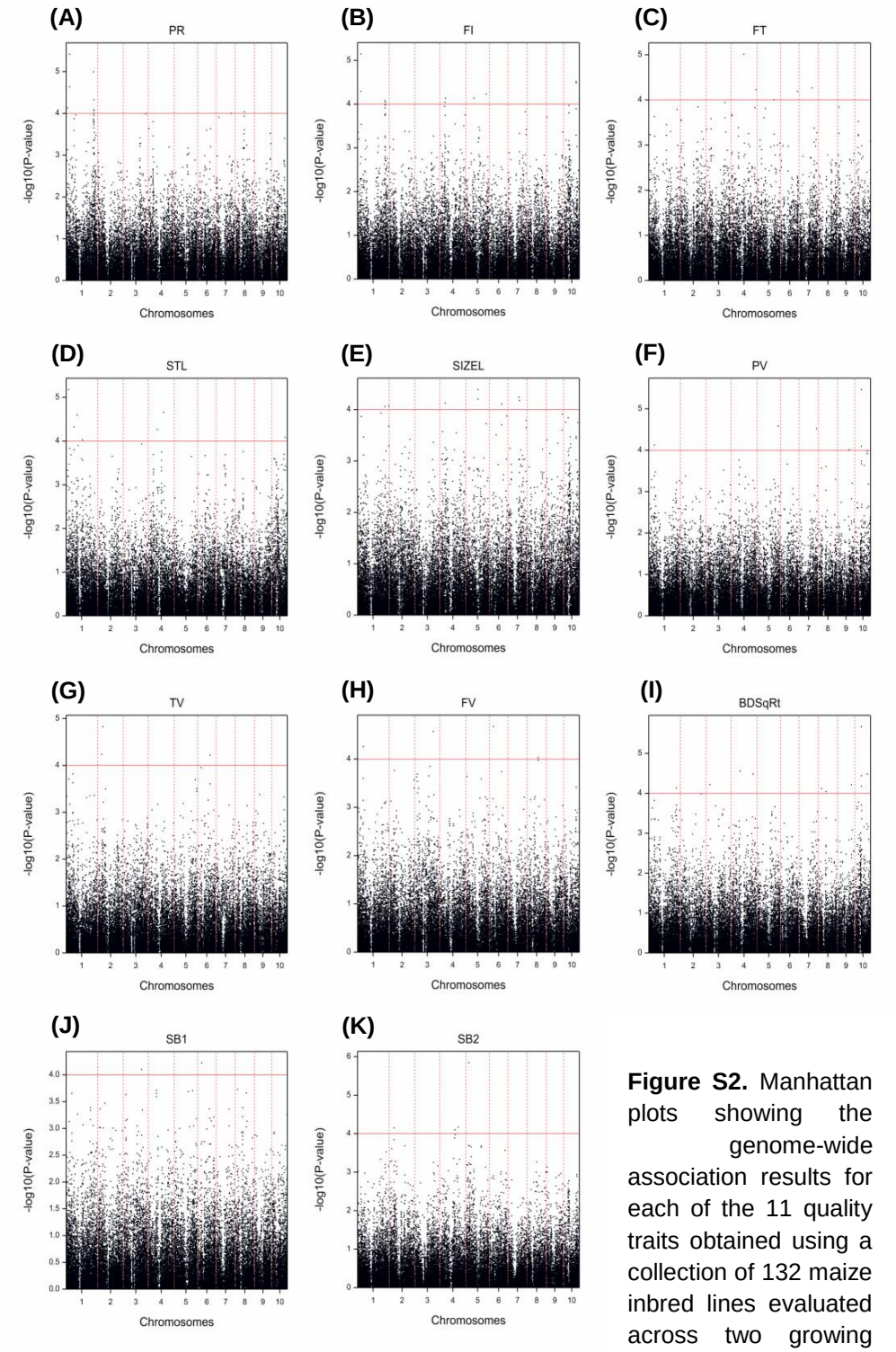

Figure S2. Manhattan plots showing the genome-wide association results for each of the 11 quality traits obtained using a collection of 132 maize inbred lines evaluated across two growing seasons. **(A)** PR – percentage of protein; **(B)** FI – percentage of fiber; **(C)** FT

– percentage of fat; **(D)** STL – percentage of starch in lyophilized flour; **(E)** SIZEL – mean particle size in lyophilized flour ; **(F)** PV – peak (maximum) viscosity; **(G)** TV – trough (minimum) viscosity; **(H)** FV – final viscosity; **(I)** BD_SqRt – squared-root transformed values of the breakdown viscosity; **(J)** SB1 – setback from trough viscosity; **(K)** SB2 – setback from peak viscosity. The *y*-axis shows the $-\log_{10}$ P values of 48,772 SNPs, and the *x*-axis shows their chromosomal positions. The horizontal red lines represent the liberal threshold of P = 1 × 10−4. Vertical lines separate each of the 10 maize chromosomes.

Chapter V

Genetic basis of carotenoids, tocopherols, and phenolic compounds in wholemeal maize flour − a genome-wide association approach

The work presented in this chapter corresponds to the following manuscript in preparation:

Alves M. L., Bento-Silva A., Carbas B., Gaspar D., Paulo M., Brites C., Mendes-Moreira P., Brites C., Bronze M. R., Malosetti M., van Eeuwijk F., & Vaz Patto M. C. Genetic basis of carotenoids, tocopherols, and phenolic compounds in wholemeal maize flour − a genome-wide association approach (*in preparation*)

In this research paper, Mara Lisa Alves performed the DNA isolation, the genotypic and phenotypic data analysis, the association mapping analysis and follow-up analysis, and drafted the manuscript. (See Acknowledgements section for authors' contributions)

Abstract

Maize is one of the most import food crops worldwide. Consumers worldwide are increasingly concerned with food quality. However, maize breeding for improved quality is a complex task and therefore developing tools to select for better quality products is of great importance nowadays. The exploitation of maize natural variation in compounds with antioxidant activity, such as tocopherols, carotenoids and phenolic compounds, has received an increased interest in the last years due to its beneficial role in human health and also due to their effect on preventing quality deterioration of food products helping to maintain their nutritional value.

In this work we took advantage, for the first time, of a collection of 132 maize inbred lines, partially developed from Portuguese traditional maize populations, to carry out a genome-wide association study aiming to identify genomic regions controlling for several antioxidant compounds of maize wholemeal flour, fundamental on the development of quality-related molecular selection tools. The inbred lines were trialed during two growing seasons and samples from each field replicate characterized for total carotenoids content, α-tocopherol, γ-tocopherol, δ-tocopherol, total free phenolic compounds and hydroxycinnamic acids content, using well-established methodologies. Each maize inbred line was previously genotyped with the MaizeSNP50 BeadChip array. SNP-trait associations were tested using a mixed linear model accounting for the genetic relatedness among inbred lines.

With this approach, 73 different genomic regions were identified associated with the 10 antioxidant compounds-related traits evaluated. The majority of the identified genomic regions were associated to a single trait (78%). The stronger SNPs associations

with trait variation were detected for total carotenoids content, flour yellowness, and lightness, on chromosome 6, for δ-tocopherol content, on chromosome 1, and α-tocopherol content, on chromosome 5. Several of the SNP-trait associations were located within or near genes known to be involved in the carotenoids and tocopherols biosynthetic pathway. The strongest SNP-trait associations for total carotenoids content, flour yellowness, and flour lightness were located upstream of the GRMZM2G300348 gene (*y1 - yellow endosperm1*), coding for a phytoene synthase (PSY1), an enzyme catalyzing the first committed step of the carotenoids biosynthetic pathway.

Although for all the traits analyzed significant SNP-trait association were detected, this study was particularly successful in unveiling the genetic architecture of traits either with high heritability values, controlled by a smaller set of genes, and/or traits controlled by large-effect *loci* (e.g., flour yellowness and total carotenoids content).

Keywords: *Zea mays* L., carotenoids, tocopherols, phenolic compounds, color, Portuguese maize germplasm, quantitative trait loci, candidate genes

1 Introduction

In the last few decades, consumers' views on how food positively or negatively affects their health have changed.Today, food is not only intended to satisfy hunger and provide the necessary nutrients, but it is also used to prevent nutrition-related diseases and improve physical and mental well-being (reviewed in Siró et al., 2008). Maize (*Zea mays* L.) is, along with rice and wheat, one of the world leading crops and a crucial source of food, feed, fuel and fibers

(Tenaillon & Charcosset, 2011). Together, these three species account for 93% of all cereal food consumption (FAO, 2012), playing a major role in nutrition in many countries, as a source of oil, flour and starch (Fernandes et al., 2013). Besides maize kernel content in macronutrients, such as starch, protein, oil, and fiber (reviewed in Chen et al., 2016a), maize is also rich in micronutrients such as several vitamins (Nuss & Tanumihardjo, 2010) and phenolic compounds (Bento-Silva et al., 2017) that may contribute to their overall antioxidant activity. The exploitation of maize natural variation in compounds with antioxidant activity, such as tocopherols, carotenoids and phenolic compounds, has received an increased interest in the last years. This interest can be partially explained by its benefits for human health in the prevention of chronic diseases (Ktenioudaki et al., 2015), but also due to their effect on the prevention of quality deterioration of food products, contributing to the maintenance of their nutritional value (Shahidi, 1997).

In maize kernel, the highest amount of phenolic compounds is present in the insoluble fraction, but the soluble/free fraction has high chemical diversity, being strongly associated with the color of the kernel (Salinas-Moreno et al., 2017). Moreover, the presence of those compounds in different plant-based foods may contribute to food flavor and color (Salinas-Moreno et al., 2017). Vitamin A, as provitamin A carotenoids, and vitamin E, as tocopherols, are the predominant fat-soluble vitamins found in maize kernels (Nuss & Tanumihardjo, 2010). Carotenoid accumulation conveys a yellow-orange color to the endosperm (Wurtzel et al., 2012). Vitamin E is found almost exclusively in maize germ oil at about 94% of total tocopherols (reviewed in Nuss & Tanumihardjo, 2010). Natural genetic variability in carotenoids, tocopherols, and phenolic compounds has been reported in maize (e.g., Žilić et al., 2012).

Given the rising awareness in consumers' dietary choices, the consideration of quality aspects in plant breeding is now a commercially relevant issue. The diversity of maize has been the base for breeding programs that have generated much of the higher-yielding maize varieties presently used worldwide. Historically, this effort has primarily focused on increasing stability and grain yield potential, under abiotic and biotic stresses (reviewed in Muzhingi et al., 2017). In the last decade, however, much effort has been made in evaluating and using the diversity of maize also on the improvement of animal feed and human nutrition (reviewed in Muzhingi et al., 2017). As reviewed by Moose & Mumm (2008), conventional plant breeding that relies only on phenotypic selection has been historically effective. However, for some traits, phenotypic selection has made little progress due to challenges in measuring phenotypes accurately or in the identification of the individuals with the highest breeding value. The effects of environment, genotype-by-environment interaction, and measurement errors also contribute to the reduced progress. For some traits, only destructive measurements are available to accurately access the phenotype, or trait expression may be dependent on the developmental stage (e.g. kernel quality traits) (Moose & Mumm, 2008). Recently, Wen et al. (2016a) and Jiang et al. (2017) reviewed the advances made in maize improvement. These authors concluded that currently, the efforts for improving maize kernel covers not only the traditional staples of oil, protein, and starch traits but also compounds such as vitamins and free amino acids contents, as well as secondary metabolites such as phenylpropanoids and alkaloids.

Maize is the basis for the production of several foods, such as polenta, bread, tortillas, snacks, cornflakes (Fernandes et al., 2013). In some countries, such as Spain and Portugal, wholemeal maize

flour is used for bread production (Brites et al., 2010; Rodríguez et al., 2013). The Portuguese maize germplasm is recognized by its high diversity (Vaz Patto et al., 2004, 2007, 2009) and associated potential quality for food since Portugal has a long tradition in the production of an ethnic leavened maize-based bread – *broa* (Vaz Patto et al., 2007). This ethnic bread is made with a 50% or more of maize flour, mixed with wheat or rye (Brites et al., 2010), for which the local maize populations are usually preferred (Vaz Patto et al., 2007). The fact that flour produced from locally grown maize populations has been traditionally preferred in the formulation of *broa* has been pointed out by Vaz Patto et al. (2007) as one of the reasons for the present on-farm survival of the Portuguese maize populations. All in all, bread making requires a deep understanding of the many complex raw material and process interactions that collectively contribute to the final product quality (Cauvain, 2012). Maize flour from the Portuguese maize populations have, on average, higher levels of protein and fiber and lower levels of α- and δ-tocopherols, associated with a lower breakdown of viscosities values when compared to maize populations from other origins (Alves et al., Chapter III). In the same study, it was verified that on a quality-oriented maize breeding program using the Portuguese populations, breeding objectives should focus on increasing the agronomic performance of the populations but also on their tocopherol levels since these are limiting on this germplasm (Alves et al., Chapter III). An increase in maize vitamin E levels, as tocopherols, would elevate maize nutritional value (Nuss & Tanumihardjo, 2010).

Variability on the ferulic acid and *p*-coumaric acid content, the two main phenolic compounds found in maize kernel (Adom & Liu, 2002; Pei et al., 2016), and on the total carotenoid content was previously reported among Portuguese traditional maize populations

(Alves et al., Chapter III). It has been shown that some of these antioxidant compounds may reduce the retrogradation and improve starch qualities (Beta & Corke, 2004; Siriamornpun et al., 2016; Zhu et al., 2009), or influence the formation of dough texture (Klepacka & Fornal, 2006), a very important parameter in defining bread quality (Matos & Rosell, 2012). Additionally, secondary metabolites such as carotenoids, but also phenolic compounds are known to greatly contribute to maize kernel color (Žilić et al., 2012). Kernel color is generally linked to consumer acceptance (Ranum et al., 2014) and appears also to be important for Portuguese maize bread consumer choices (Carbas et al., 2016).

Most agriculturally and economically important traits have complex genetic underpinnings (i.e., determined by multiple quantitative trait loci, QTLs) (Wen et al., 2016a). Precise location and characterization of these functional loci can facilitate crop improvement *via* marker-assisted selection (MAS). To dissect complex traits, linkage analysis and association mapping have been commonly used (Wen et al., 2016a). The underlying genetic basis for the variation on tocopherol and carotenoids levels in maize kernel, and to a less extend on phenolic compounds variation, has been the subject of quantitative trait locus (QTL) linkage mapping, and during last years this research was boosted by association analysis (Wong et al., 2004; Chander et al., 2008a; Harjes et al., 2008; Yan et al., 2010; Li et al., 2012; Shutu et al, 2012; Azmach et al., 2013; Chandler et al., 2013; Fu et al., 2013; Lipka et al., 2013; Romay et al., 2013; Owens et al., 2014; Wen et al., 2014; Suwarno et al., 2015; Santiago et al., 2016; Wen et al., 2016b; Diepenbrock et al., 2017; Jittham et al., 2017). As pointed out by several authors (e.g., Shutu et al., 2012; Treutter, 2010; Zhai et al., 2016), for devising more efficient breeding tools to support the improvement of these health-promoting

compounds, a comprehensive and deeper understanding of the regulatory mechanisms and the complex genetic basis of maize flour antioxidants is essential.

A comprehensive analysis of all these different quality-related parameters is still missing in the Portuguese maize germplasm and so the national diversity was never properly exploited on quality breeding neither on the development of efficient tools / innovative approaches to support breeding for these complex quality traits. In this work, we took advantage of the diverse germplasm developed through decades of maize breeding by several Portuguese regional maize breeding stations now extinct such as the NUMI (Núcleo de Melhoramento de Milho), and presently conserved at the Portuguese Bank of Plant Germplasm (Banco Português de Germoplasma Vegetal - BPGV, Braga, Portugal). This collection of maize inbred lines, containing a considerable amount of the unexplored Portuguese maize germplasm, was characterized for the following antioxidant compounds-related traits, measured in wholemeal maize flour: total carotenoids, α-tocopherol, γ-tocopherol, δ-tocopherol, total free phenolic compounds and total hydroxycinnamic acids content. Different flour color parameters were also measured: flour lightness, red/green index, and yellowness. The main objective was to identify genomic regions controlling the upper mentioned quality-related parameters through a genome-wide association approach. The uniqueness of the association mapping panel used in the current work, constituted by Portuguese, foreign and mixed origin lines, could lead to the discovery of genomic regions associated to the variation of quality traits not previously identified in other germplasm collections analysis.

2 Materials and Methods

2.1 Plant material

The maize inbred line collection used in this study was the same as previously used in Alves et al. (unpublished, Chapter IV). As described previously in Chapter IV, from a total of 164 different maize inbred lines sowed on the field trials, only 132 yielded sufficient kernels to proceed with their quality analysis (Table S1). Additional details on their recorded pedigree may be found also in Table S1. This collection varied in kernel color from white, yellow, yellow-orange, orange, and red, and in endosperm type from flint, intermediate, and dent types. The summary showing those lines grouped by endosperm type and kernel color can be found in Chapter IV, Material and Methods section (see Table 1 in Chapter IV).

This collection was assembled observing a significant representation of lines selected from traditional Portuguese maize populations (29 lines) and lines with a mixed Portuguese x foreign origin (the majority of the lines whose names start by *PB*, *PP*, *PV* or *PT*, Table S1). The rationale behind this was the premise that the locally grown Portuguese maize populations, is the material traditionally used for the formulation of food commodities, were considered as keepers of quality traits related to food production. The original seed of the maize inbred lines collection used in this study was provided by the Portuguese Bank of Plant Germplasm (BPGV, Braga, Portugal).

2.2 Field characterization and experimental design

The inbred lines were evaluated at Coimbra site (40°13'0.22"N, 8°26'47.69"W), Portugal, during the 2011 and 2012 growing seasons, using an organic agriculture converted field. The site characterization

and experimental design were previously described in Chapter IV. Briefly, in each year, the maize inbred lines were evaluated using a randomized complete block design, with two blocks (replicates). Information on the spatial distribution of the plots was also recorded (row and columns field coordinates). Each plot consisted of two rows 7.2 m long (6.4 m planted row plus 0.8 m border space between two planted rows), with an inter-row distance of 0.75 m. Plots were mechanically and hand-weeded when needed and managed following common agricultural practices for maize in the region. Pollination was controlled within each plot. All the plots were harvested by hand.

After harvest, ears were dried at 30-35ºC in an oven (Memmert Model UFE 800, Memmert GmbH + Co. KG, Germany) until a ~15% moisture was reached. The ears were then shelled and the kernel collected per plot basis, packed in paper bags and kept at 4ºC until further analysis.

2.3 Phenotypic data acquisition

A seed sample from each of the harvested plots (replicates) was used for quality determinations. The total number of samples analyzed corresponded therefore to [*number of inbred lines* × *number of field replicates (2)* × *number of growing seasons (2)*].

Firstly, wholemeal maize flour was obtained from all the seed samples using a Falling number 3100 mill (Perten Inc., Sweden) with a 0.8 mm screen.

In this work, the following antioxidant compounds-related traits, measured in this maize flour, were considered: total carotenoids content (TCC), α-tocopherol (AT), γ-tocopherol (GT), δ-tocopherol (DT), total phenolic compounds content as assessed by *Folin–Ciocalteu* assay (PHS) and by HPLC (PHH), total hydroxycinnamic acids content (HY). Additionally, several color parameters were

measured: flour lightness (*L**), red/green index (*a**), and yellowness (*b**).

2.3.1. Total carotenoids content

The total carotenoids content (TCC) was spectrophotometrically measured at 450 nm according to the AACC method 14-60.01 (AACC International, 2012). Results were expressed in µgrams of lutein equivalent per gram of sample, as the main carotenoid found in maize.

2.3.2 Tocopherols content

α-Tocopherol (AT), γ-tocopherol (GT), δ-tocopherol (DT) were separated from the fat portion of the maize flours by high-performance liquid chromatography (HPLC) and quantified using an Agilent 1200 model with a fluorescence detector (FLD) and a Diol column (LiChropher 100, 250 x 4 mm) according to the method ISO 9936 (2006). Tocopherols content was expressed in µg/g fat basis.

2.3.3 Total phenolic compounds content

Ethanolic extracts (EtOH:H_2O 50:50, v/v) for assessing the total phenolic content (PH) of maize flour were prepared following the procedure described by Lopez-Martinez et al. (2009), with some modifications. Briefly, 2 g of maize flour were extracted with 20 mL of EtOH:H_2O (50:50, v/v) for 15 minutes, using an Ultra Turrax T25 (Janke & Kunkel, IKA Labortechnik, Germany). Final extracts were filtered using a Whatman filter paper (type42: retention 2.5 µm, diameter 18.5 cm). Extracts were prepared in triplicate and preserved at -20ºC until analysis.

Total phenolic compounds content was assessed using two different methodologies: by *Folin-Ciocalteau* assay (Singleton et al., 1999), and by high-performance liquid chromatography (HPLC). The

reasoning behind using both methods was that the *Folin–Ciocalteu* assay has historically been the most used methodology for measuring the total phenolic compounds content. However, as reviewed by Naczk and Shahidi (2006), the quantifications by HPLC are more reliable as the measurements obtained from the *Folin-Ciocalteau* assay can suffer from an overestimation due to the presence of other compounds that absorb in the same wavelength used in the assay. The levels of total hydroxycinnamic acids in each sample were also assessed by HPLC.

The *Folin-Ciocalteau* assay (Singleton et al., 1999) used to determine the total phenolic compounds (PHS) content was performed using a Beckman DU-70 spectrophotometer, with slight modifications as described in Silva et al. (2015), and the total free phenolic compounds content per sample was expressed in mg of gallic acid equivalents/100 g of dry weight (GAE/100 g DW).

Total phenolic compounds (PHH) and total hydroxycinnamic acids (HY) content were quantified by High-Performance Liquid Chromatography (HPLC). The HPLC system used was a Thermo Finnigan (Surveyor model) equipped with an autosampler, pump and photodiode array detector (PDA) coupled to a Dionex ED40 electrochemical detector. Chromatographic separation of compounds was carried out with a Lichrocart RP-18 column (250 x 4 mm, particle size 5 µm, Merck) and a Manu-cart® RP-18 pre-column in a thermostated oven at 35°C.

Photodiode array detector was programmed for a scanning between 192 and 798 nm at a speed of 1 Hz with a bandwidth of 5 nm. The detection was monitored using three individual channels, 280, 320 and 360 nm, at a speed of 10 Hz with a bandwidth of 11 nm. The injection volume applied was 20 µL.

The auto sampler's temperature was set at 12°C. The eluents used were A – phosphoric acid solution p.a. (0.1%) in Milli-Q® water and B – 0.1% phosphoric acid p.a. in acetonitrile HPLC gradient grade: Milli-Q® water (40:59.9), at a flow rate of 0.700 mL/min. The following gradient of eluents was used: initially 100% A and 0% B, 0-20% B over 15 min, 20% B for 10 min, 20-70% B from 25 to 70 min, 70% B for 5 min, 70-100% B from 75 to 85 min, held isocratically (100% B) for 15 min, followed by a equilibration step of 10 min.

The total phenolic content was determined by total chromatogram areas at 280 nm. Results were expressed in mg gallic acid equivalents (GAE) / 100 g of dry weight.

The total hydroxycinnamic acid content was determined by the total chromatogram areas of hydroxycinnamic acids compounds at 320 nm. Results were expressed in mg ferulic acid equivalents (FAE) / 100 g of dry weight.

2.3.4 Flour color parameters

Maize flour color was assessed on a 10 to 12 g sample in an opaque recipient using a Minolta chromameter CR-2b with the CIE tristimulus color parameters: *L** - lightness, *a** - red/green index, *b** - yellow/blue index. *L** values can vary from *L** = 0 (black) to *L** = 100 (white); positive *a** values meant that samples tend toward the red part of the color spectra; positive *b** values meant that samples tend toward the yellow part of the color spectra.

2.4 Phenotypic data analysis

A phenotypic data analysis was performed per individual trial, as already described in Chapter IV (Alves et al., unpublished) to 1) perform quality control of the data, 2) obtain estimates of genetic variances (and covariances between traits) and heritabilities, and 3)

obtain adjusted trait means per inbred line. The phenolic compounds (PHS, PHH, and HY), and α- and δ-tocopherol (AT and DT) data required a $\log_{10}$-transformation to stabilize the variance. All analyses were performed, following the same procedure as described in Chapter IV (Alves et al., unpublished), using the Breeding View software (Murray et al., 2014), available through the IBP Breeding Management System (The IBP Breeding Management System Version 3.0.9, 2015).

Traits' heritability, as well as the best linear unbiased predictors (BLUPs) for each inbred line, was calculated. The correlations between BLUPs of different traits were used to obtain estimates of genetic correlations between traits. The adjusted means for field trial design (best linear unbiased estimators - BLUEs) for each growing season were retrieved to be used afterward as the input phenotypic data on the association mapping analysis and to assess the phenotypic correlations between traits.

For each trait, a multi-environment trial analysis was also performed to test for interaction between the maize inbred lines and the two growing seasons, as previously described in Chapter IV (Alves et al., unpublished). The analysis of variance was carried out using the REML variance components analysis procedure in Genstat software (Genstat® for Windows 18th edition, Payne et al., 2015). The mixed model included growing seasons (fixed), maize inbred lines and season by line interaction (fixed or random), while blocks, rows, and columns, were treated as random terms, and nested within growing seasons. BLUPs and BLUEs were calculated for each inbred line across growing seasons. BLUPs were used on principal component analysis (PCA) to assess genetic correlations between traits and BLUEs were used as input phenotypic data in the

association mapping analysis, for the combined analysis across growing seasons.

2.5 Genotypic data

The same genotypic dataset as previously described in Chapter IV (Alves et al., unpublished) was used to perform the association analysis. Briefly, each maize inbred line was genotyped with the Illumina MaizeSNP50 BeadChip array (Ganal et al., 2011) using genomic DNA obtained from adult leaves. The genotyping array procedure and alleles scoring was conducted by the genotypic service provider (TraitGenetics GmbH, Gatersleben, Germany). This array allows the screening of 17,520 genes (since 33,417 of the SNPs present in this array are located on 17,520 genes and 16,168 SNPs are located in intergenic regions) (Ganal et al., 2011). The position of each marker along the maize B73 reference genome was updated from the initially available coordinates when the MaizeSNP50 BeadChip was originally designed (B73 reference genome version 1) to the coordinates in the released B73 reference genome version 3. These coordinates were taken from the maize genome browser, via the MaizeGDB database (Lawrence et al., 2008, www.maizegdb.org). Genotypic data quality control was performed by removing SNP markers and inbred lines with more than 25% of missing data. SNPs called as heterozygous were set as missing data (0.93% of the total SNP calls). Moreover, markers close to fixation (allelic frequency superior to 95%) or markers with a minor allele frequency (MAF) smaller than 5% were also removed. After this filter, a total of 48,772 SNPs remained and were used for the association mapping analysis.

2.7 Association mapping analysis

Given that for all the quality traits under study, with the exception of δ-tocopherol (DTLOG), variance components for genotype-by-environment (G × E) interaction ($\sigma^2_{g\times y}$) were much smaller than the genotype main effect variance component ($\sigma^2 g$), univariate association analysis was carried out using the adjusted means from the field trial design (BLUEs) obtained across growing seasons. Additionally, and only for δ-tocopherol (DTLOG), the univariate association analysis was also carried out using the adjusted means (BLUEs) obtained separately for each growing season. Genome-wide association studies were conducted with the Genstat software using the available genotypic (SNPs from the MaizeSNP50 BeadChip array) and antioxidant compounds-related data (10 quality traits) measured in 132 maize inbred lines. The Genstat software performs association mapping in the mixed model framework, fitting markers as fixed and inbred lines as random terms using REML (Malosetti et al., 2007).

Following the same procedure described in Chapter IV (Alves et al., unpublished), three different models were tested to detect significant marker-trait associations: the naïve model [Phenotype = SNP + (Genotype + Error)], that neither accounts for population structure nor for familiar relatedness; a model accounting for population structure (Q) using 15 principal components from PCA [Phenotype = Q + SNP + (Genotype + Error)]; and a model accounting for familiar relatedness (K) [Phenotype = SNP + Genotype + Error] with Genotype random effects structured following a kinship matrix K. For each chromosome, a different kinship matrix was calculated where only the SNPs located on the other nine maize chromosomes were used to calculate the kinship matrix (Listgarten et

al., 2012; Rincent et al., 2014). The principal components to account for population structure among inbred lines and the kinship matrix to account for pairwise genetic relatedness among inbred lines were previously calculated in Alves et al. (unpublished, Chapter IV) using a subset of 1,821 SNPs, evenly distributed across the genome (corresponding approximately to 1 SNP per Megabase pairs).

The observed P-values from marker-trait associations were used to draw Manhattan plots. The threshold to consider a marker-trait association significant was set to $-\log_{10}$ (P-value = 4). Additionally, the more stringent Bonferroni-corrected thresholds at $\alpha = 0.05$ (P-value = 1/N, N represents the number of markers used in GWAS for each chromosome) worked as the threshold-guideline to discuss the strongest association detected for each trait. The effect of the minor frequency SNP variant, reported in relation to the most frequent allele reference, was calculated.

2.8 Post-GWAS procedures

To define the chromosomal regions where to search for candidate genes for the traits under analysis, a local linkage disequilibrium (LD) study was performed following the same approach as already described in Chapter IV (Alves et al., unpublished). LD was estimated as the squared correlation coefficient, r^2, after correcting for population structure using the principal component scores from Eigenanalysis, as implemented in Genstat software.

The markers' positions flanking each local LD block, for which LD $r^2 > 0.2$, were further used as queries positions on the maize genome browser, via MaizeGDB (https://www.maizegdb.org/gbrowse), to retrieve the list of candidate genes mapped within those genomic regions. The genome sequence of the maize inbred line B73 (*Zea mays* B73 RefGen_v3 assembly)

was used as the reference genome for candidate gene analyses (Schnable et al., 2009). The functional annotation of the genes under the genomic regions identified in the GWAS was retrieved via Phytozome (Goodstein et al., 2011, Phytozome 11, version AGPv3 - Zea mays Ensembl-18) using the gene model identifier as the query. KEGG: Kyoto Encyclopedia of Genes and Genomes database (Kanehisa & Goto, 2000) was used to retrieve information on the pathways where the candidate genes could be involved.

3 Results

3.1 Wholemeal maize flour antioxidant compounds-related traits variation

As shown in Table 1, where the quality traits variance components and heritabilities are presented, the highest percentage of variance was normally due to differences between the inbred lines (σ^2g), except for PHSLOG, where the error variance component was higher than the genotype main effect variance, and DTLOG, where the σ^2g×y variance component was higher than the genotype main effect variance. Nevertheless, with the exception of DTLOG, for all the other traits analyzed, the variance component associated to differences between inbred lines was far greater than the variance component attributed to the G×E interaction term effect (σ^2g / σ^2g×y > 1).

Total carotenoids content (TCC) and flour yellowness (*b**) had by far the highest heritability value across growing seasons (h^2 = 96%, for *b**, and h^2 = 92%, for TCC). By contrast, DTLOG and PHSLOG had the lowest heritability values (h^2 = 37%, for DTLOG, and h^2 = 41%, for PHSLOG) (Table 1). Complementary information on the collection of maize inbred lines phenotypic values (range, and

mean ± standard deviation) for the quality traits evaluated during two growing seasons (2011 and 2012) can be found in Table S4.

Table 1. Variance components and broad sense heritability values for quality traits in wholemeal flour of a collection of 132 maize inbred lines evaluated in two growing seasons (2011 and 2012).

Trait	Variance components [4]							h^2 heritability [5](%)		
	σ^2p	σ^2g	$\sigma^2g\times y$	σ^2error	σ^2g (%)	$\sigma^2g\times y$ (%)	σ^2error (%)	2011	2012	Across growing seasons
*L**, in CIE color system units	3.06	2.36	0.26	0.43	77	9	14	63	72	77
*a**, in CIE color system units	0.50	0.39	0.05	0.06	79	10	11	65	74	79
*b**, in CIE color system units	69.81	67.21	1.23	1.36	96	2	2	75	70	96
TCC, in µg lutein equivalents/g of sample	566.72	522.69	32.50	11.52	92	6	2	71	75	92
ATLOG[1], in µg/g fat	0.37	0.22	0.11	0.04	59	31	10	62	75	59
DTLOG[1], in µg/g fat	0.16	0.06	0.08	0.02	37	53	10	67	72	37
GT, in µg/g fat	20.11×10^{3}	13.93×10^{3}	1.37×10^{3}	4.81×10^{3}	69	7	24	61	69	69
PHSLOG[1 2], in mg GAE/100 g DW	1.00×10^{-2}	0.30×10^{-2}	0.10×10^{-2}	0.50×10^{-2}	41	8	51	41	58	41
PHHLOG[1 2], in mg GAE/100 g DW	1.21×10^{-2}	0.79×10^{-2}	0.03×10^{-2}	0.40×10^{-2}	65	2	33	53	65	65
HYLOG[1 3], in mg FAE/100 g DW	13.7×10^{-2}	8.3×10^{-2}	0.70×10^{-2}	4.6×10^{-2}	61	5	34	52	66	61

[1] Log10 transformed values | [2] mg GAE / 100 g DW – milligrams of gallic acid equivalents per 100 grams of dry weight | [3] mg FAE / 100 g DW - milligrams ferulic acid equivalents per 100 grams of dry weight | [4] Percentage of variance attributed to the individual terms of the statistical model: σ^2p corresponds to the phenotypic variance; σ^2g corresponds to the genotypic variance; $\sigma^2g\times y$ corresponds to the interaction between inbred lines and growing seasons variance; σ^2error (%) corresponds to the error variance; σ^2g (%) corresponds to the percentage of variance attributed to difference among inbred lines; $\sigma^2g\times y$ (%) corresponds to the percentage of variance attributed to the interaction between inbred lines and growing seasons; σ^2error (%) corresponds to the percentage of variance attributed to the block, row, column, and residual terms which altogether compose the error variance | [5] h^2=broad sense heritability obtained by fitting inbred lines as random terms in the statistical model. | Quality traits' abbreviations: L - flour's lightness; a* - flour's red/green index; b* - flour's yellow/blue index; TCC - total carotenoids content; ATLOG - α-tocopherol content; DTLOG - δ-tocopherol content; GT - γ-tocopherol content; PHSLOG - total phenolic compounds by Folin–Ciocalteu assay; PHHLOG - total phenolic compounds by HPLC; HYLOG - total hydroxycinnamic acids*

Considering the data obtained across the two growing seasons, flour yellowness (*b**) was highly and positively correlated with the level of total carotenoids (TCC) content on the samples (r> 0.8). Flour lightness (*L**) was moderately and negatively correlated with *a**, *b** and TCC (r < −0.4). The assessment of the total phenolic compounds content measured with the *Folin-Ciocalteau* assay and HPLC were positively correlated (r > 0.6). HYLOG was moderately and positively correlated with the total phenolic compounds as measured by HPLC (r > 0.5) (Figure 1, Table S2 and Table S3).

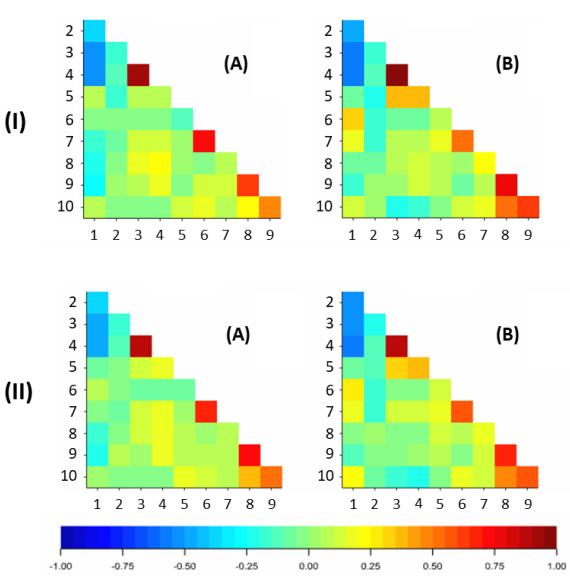

Figure 1. Heat maps illustrating the (I) phenotypic and (II) genetic correlations for health-related and color quality traits measured in wholemeal flour of 132 maize inbred lines grown during (A) 2011 growing season, and (B) 2012 growing season. Quality traits' key: 1 – *L**, flour's lightness; 2 – *a**, flour's red/green index, positive values indicate that samples tend toward the red part of the color spectra; 3 – *b**, flour's yellow/blue index, positive values indicates that samples tend toward the yellow part of the color spectra; 4 – TCC, total carotenoids content; 5 – ATLOG, $\log_{10}$-transformed α-tocopherol content; 6 – DTLOG, $\log_{10}$-transformed δ-tocopherol content; 7 – GT, γ-tocopherol content; 8 – PHSLOG, $\log_{10}$-transformed total phenolic compounds, assessed by Folin–Ciocalteu assay; 9 – PHHLOG, $\log_{10}$-transformed total phenolic compounds, assessed by HPLC; 10 – HYLOG, $\log_{10}$-transformed total hydroxycinnamic acids, assessed by HPLC.

In Figure 2, the first two components of the principal component analysis, explaining a total of 48.43% of the variability present in the dataset, are projected depicting a high diversity among the inbred lines of the association panel for the majority of the traits analyzed.

The maize inbred lines could clearly be separated into two main groups according to their flour yellowness and total carotenoids content. Representatives from the inbred lines mainly derived from Portuguese traditional maize populations could be found in both color groups.

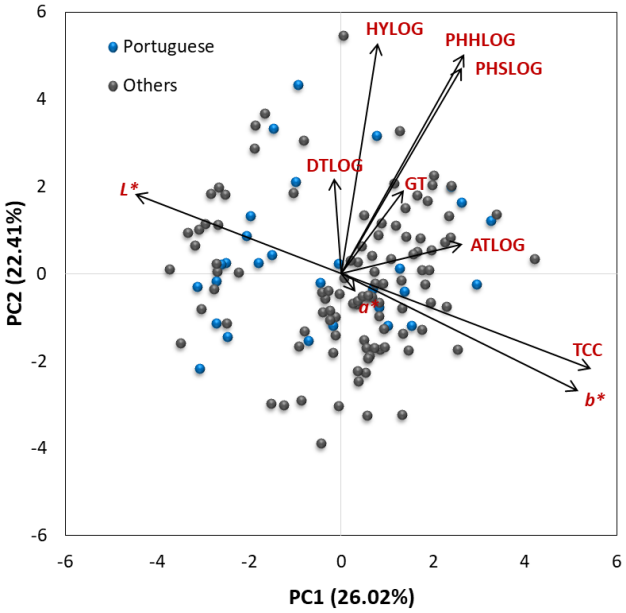

Figure 2. Principal component analysis (PCA) biplot based on BLUP values for 11 quality traits measured in 132 maize inbred lines. Circles colored in blue correspond to inbred lines selected entirely from Portuguese landraces. Quality traits' abbreviations: *L** – flour's lightness; *a** – flour's red/green index, positive values indicate that samples tend toward the red part of the color spectra; *b** – flour's yellow/blue index, positive values indicates that samples tend toward the yellow part of the color spectra; TCC – total carotenoids content; ATLOG – log_{10}-transformed α-tocopherol content; DTLOG – log_{10}-transformed δ-tocopherol content; GT – γ-tocopherol content; PHSLOG – log_{10}-transformed total phenolic compounds, assessed by Folin–Ciocalteu assay; PHHLOG –

log_{10}-transformed total phenolic compounds, assessed by HPLC; HYLOG – log_{10}-transformed total hydroxycinnamic acids, assessed by HPLC.

3.2 Genomic regions associated with quality traits

GWAS was performed using a mixed linear model (MLM) and either kinship relationship (K matrix) or population structure (Eigenanalysis) was taken into account to avoid spurious associations. After inspecting the observed inflation factors obtained for each tested model, the mixed linear model accounting for familial relatedness (K matrix) was selected as the best model (Table S5). Therefore, the results reported below concern the results obtained using this model.

For all the studied tocopherols, carotenoids and phenolic compounds, significantly associated SNP markers were identified (Figure 3). In total, and using a liberal threshold $-log_{10}$(P-value) = 4 to recognize a significant SNP-trait associations, 104 unique SNPs were identified as being associated with the 10 quality traits analyzed across the two growing seasons (2011 and 2012), corresponding to 73 genomic regions (LD $r^2 > 0.2$) (Figure 3, Table S6). Based on the rare allele contributions to the trait variation, SNPs variants associated with increase as well as decrease in the trait value were detected for flour's lightness (*L**) and yellowness (*b**), and for total carotenoids content (TCC), α-tocopherol (ATLOG), δ-tocopherol (DTLOG), and γ-tocopherol (GT) (Table S7). For flour's red-green color (*a**), total hydroxycinnamic acids (HYLOG) and for total phenolic compounds measured by HPLC (PHHLOG), the rare allele was always responsible for an increase in the trait value (Table S7). For total phenolic compounds measured by the *Folin-Ciocalteau* assay (PHSLOG), the rare allele was always responsible for a decrease in the trait value (Table S7).

Some of the genomic regions were associated with multiple highly correlated traits. Flour lightness (*L**) and flour red-green index (*a**) shared a region on chromosome 9, and flour's yellowness (*b**) and total carotenoids content (TCC) shared 13 regions on chromosome 6, one region on chromosome 7, and one region on chromosome 9 (Figure 3). Nevertheless, the majority of the detected associated genomic regions were only associated to a single trait (78%, 57 genomic regions).

Considering the number of identified associated genomic regions across years per trait (Figure 3, see also supplementary material Table S6 and Figure S1), flour's yellowness (*b**) and total carotenoids content (TCC) appeared as the traits with the bigger number of detected associated regions (twenty-two and twenty-one regions, respectively), followed by flour's lightness (twelve regions). For all these three traits, the highest number of associations was detected on chromosome 6.

Flour's yellowness (*b**) and total carotenoids content (TCC) were simultaneously associated in thirteen genomic regions on chromosome 6 (Figure 3). Those thirteen regions on chromosome 6 were found in neighbor LD blocks, all close to each other, creating a major genomic region strongly associated to both traits spanning from 78,981 kb to 82,864 kb (Figure 3; LD blocks no.41 to no.53, in Table S6). In that major region, twenty-seven significant SNP- *b** and twenty-eight significant SNP-TCC associations were detected. The strongest association for both *b** and TCC traits was located on chromosome 6 around 82,179 kb (rs131576886; for *b**, $-\log_{10}$ (P-value) = 17.648, and for TCC, $-\log_{10}$ (P-value) = 12.89) (Figure 4). This SNP explained 20.7% of the phenotypic variance observed for *b**, and 17.9% of the phenotypic variance observed for TCC (Table S7). The presence of the variant allele "G" in that position led to a

decrease of (−)6.8 CIE color units in the *b** value (corresponding to a less yellow flour), and a decrease of (−)17.67 µg of lutein equivalents per 100 g of sample (Table S6). Three regions on chromosome 6 located between 82,179 kb and 82,334 kb, in addition to being associated with TCC and *b**, were also associated with flour lightness (*L**). The strongest SNP-*L** association was located on chromosome

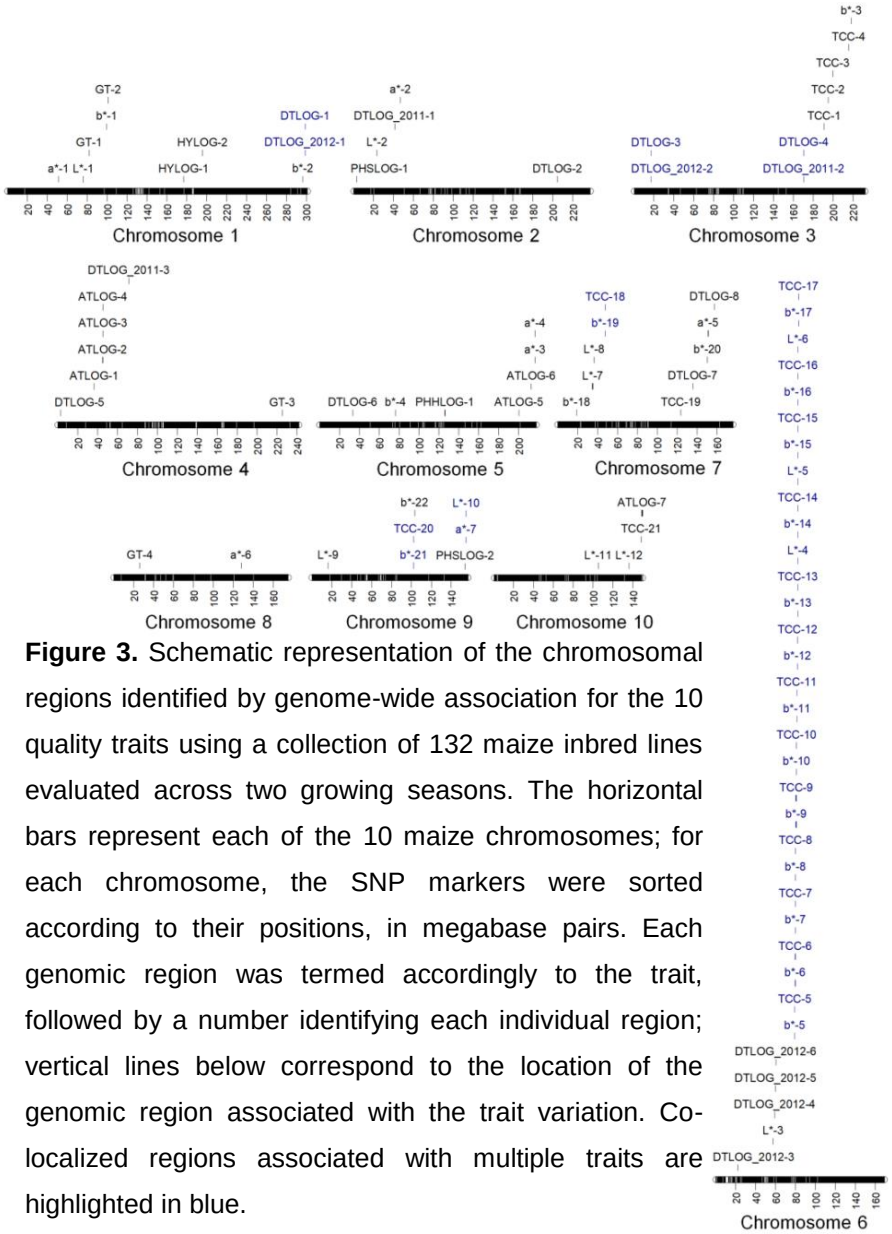

Figure 3. Schematic representation of the chromosomal regions identified by genome-wide association for the 10 quality traits using a collection of 132 maize inbred lines evaluated across two growing seasons. The horizontal bars represent each of the 10 maize chromosomes; for each chromosome, the SNP markers were sorted according to their positions, in megabase pairs. Each genomic region was termed accordingly to the trait, followed by a number identifying each individual region; vertical lines below correspond to the location of the genomic region associated with the trait variation. Co-localized regions associated with multiple traits are highlighted in blue.

Quality traits' abbreviations: *L** – flour's lightness; *a** – flour's red/green index, positive values indicate that samples tend toward the red part of the color spectra; *b** – flour's yellow/blue index, positive values indicates that samples tend toward the yellow part of the color spectra; TCC – total carotenoids content; ATLOG – log_{10}-transformed α-tocopherol content; DTLOG – log_{10}-transformed δ-tocopherol content; GT – γ-tocopherol content; PHSLOG – log_{10}-transformed total phenolic compounds, assessed by Folin–Ciocalteu assay; PHHLOG – log_{10}-transformed total phenolic compounds, assessed by HPLC; HYLOG – log_{10}-transformed total hydroxycinnamic acids, assessed by HPLC.

6 around 82,180 kb (rs131576883; $-log_{10}$ (P-values) = 6.384). This SNP explained 13.2% of the observed phenotypic variance for flour lightness. The presence of the variant allele "T" resulted in an increase of (+)1.02 CIE color units, corresponding to a lighter/brighter flour (Table S6).

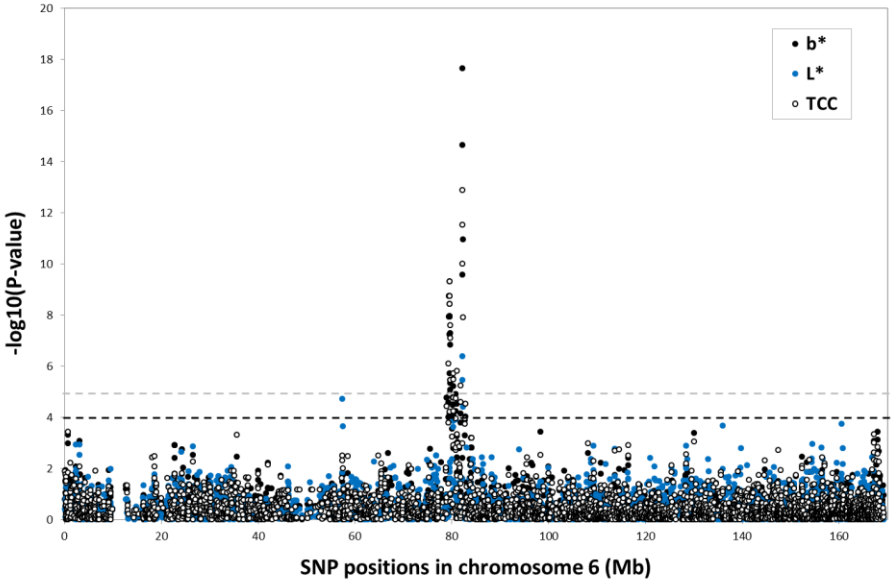

Figure 4. Chromosome 6 Manhattan plot with the genome-wide association results for flour's yellowness (*b**). Flour's lightness (*L**), and total carotenoids content (TCC) obtained using a collection of 132 maize inbred lines

evaluated across two growing seasons. The *y*-axis shows the $-\log_{10}$ *P* values of 3,922 SNPs, and the *x*-axis shows their chromosomal positions. Horizontal black and grey lines represent the liberal threshold of $P = 1 \times 10^{-4}$, and the Bonferroni-corrected threshold of $P = 1.27 \times 10^{-5}$, respectively.

As for tocopherols (α-, δ-, and γ-), δ-tocopherol (DTLOG) had the bigger number of detected associated regions (eight regions across growing seasons), plus six more regions that were only identified in the GWAS analysis done on the individual growing seasons (2011 or 2012), followed by α-tocopherol (ATLOG), with seven detected regions across chromosome 4 (four regions), 5 (two regions), and 10 (one region), and γ-tocopherol (GT) with only four regions associated located on chromosomes 1 (2 regions), 4, and 8 (Figure 3). Also for total phenolic compounds content measured by the *Folin-Ciocalteau* assay (PHSLOG) and for the total hydroxycinnamic acids content (HYLOG) only two regions were detected associated with the traits' variation, and for total phenolic compounds content measured by HPLC (PHHLOG) only one region was associated with the trait variation.

For α-tocopherol (ATLOG) the strongest SNP-ATLOG association was observed on chromosome 5 between 200,420 kb and 200,421 kb (rs130180529 and rs130180536; $-\log_{10}$ (P-values) = 5.639) (Figure 5). Both SNP were in the same LD block ($r^2 > 0.2$) and explained 13.3% of the phenotypic variance observed for α-tocopherol. The presence of the variant allele "AT" led to a reduction of (−)0.255 μgrams of α-tocopherol per gram of fat. The frequency of the variant allele "AT" on the association panel was 34.2%, however, in the 29 inbred lines derived entirely from Portuguese traditional maize populations this variant allele was the most common one (*freq* ≈ 69%).

For δ-tocopherol (DTLOG), the strongest SNP-DTLOG associations across growing seasons were located on chromosome 1 between 298,814 kb and 298,815 kb (rs128990610 and rs128990613; $-\log_{10}$ (P-value) = 5.986). Those two SNPs were positioned in the same LD block ($r^2 > 0.2$) and explained 19.66% of the phenotypic variance observed for δ-tocopherol levels across growing seasons. The presence of the "AT" variant allele led to a reduction of (−)0.272 μgrams of δ-tocopherol per gram of fat (Table S6).

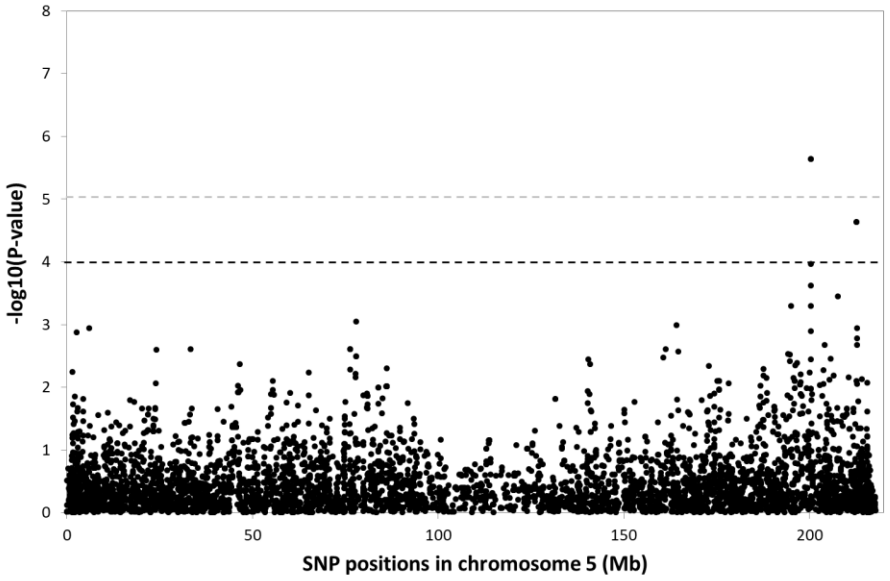

Figure 5. Chromosome 5 Manhattan plot with the genome-wide association results for flour's α-tocopherol content (ATLOG) obtained using a collection of 132 maize inbred lines evaluated across two growing seasons. The *y*-axis shows the $-\log_{10}$ *P* values of 5,305 SNPs, and the *x*-axis shows their chromosomal positions. Horizontal black and grey lines represent the liberal threshold of $P = 1 \times 10^{-4}$, and the Bonferroni-corrected threshold of $P = 9.43 \times 10^{-6}$, respectively.

For γ-tocopherol (GT), the percentage of phenotypic variance explained by each significantly associated SNP (7.22% - 9.50%) were inferior when compared with the other two tocopherols (9.58% - 13.11% for α-tocopherol, and 10.26% - 19.67% for δ-tocopherol). The strongest SNP-GT association was located on chromosome 8 around 26,077 kb (rs131179677; $-\log_{10}$ (P-value) = 4.879). This SNP explained 9.27% of the phenotypic variance observed for γ-tocopherol levels in maize flour. The presence of the "C" variant allele resulted in a reduction of (−)52.19 μgrams of γ-tocopherol per gram of fat (Table S6).

Very few significant SNP-trait associations were detected for the phenolic compound traits analyzed in this work (PHSLOG, PHHLOG, and HYLOG) (Figure 3). Moreover, the associations detected explained a small percentage of the phenotypic variance observed (6.96% - 9.57%) (Table S7). For PHSLOG, the strongest association was located on chromosome 9 at 154,122 kb (rs132575077; $-\log_{10}$ (P-values) = 4.158). This SNP explained 8% of the phenotypic variance observed. For PHHLOG, only one significant SNP-trait association was detected; it was localized on chromosome 5 between 125,841 and 126,064 kb (rs130095308; $-\log_{10}$ (P-value) = 4.214). This SNP explained 6.96% of the phenotypic variance observed. For HYLOG, the strongest association was detected on chromosome 1 at 176,867 kb (rs128811826; $-\log_{10}$ (P-value) = 4.749). This SNP explained 9.57% of the phenotypic variance observed for total hydroxycinnamic acids content.

3.3 Candidate genes identification

The average LD decay for the quality traits significantly associated genomic regions was 47.20 kb for LD $r^2 > 0.2$. This value extended to a maximum LD distance of 853 kb in a region of

chromosome 10 spanning from 148,762 kb to 149,615 kb identified as being associated with the α-tocopherol trait (ATLOG) (Table S6). Using as reference the filtered gene set from the B73 RefGen_v3 assembly, a complete list of genes mapped within the significantly associated genomic regions identified in the GWAS for the 10 quality traits can be found in Table S8. A considerable proportion of the SNPs significantly associated with the quality traits were mapped within genes (~68%, 71 out of 104 SNPs significantly associated with any trait, Table S8). And the degree of linkage disequilibrium around the genomic regions identified by GWAS allowed achieving a mapping resolution to the gene level for 54.79% of the cases (LD blocks where a single gene was identified, Table S8).

In the frame of this thesis, it is not possible to describe all candidate genes located within the associated genomic regions in detail (Table S8). We here, therefore, restrict ourselves mainly to those that were located within regions where the strongest significant associations were detected. The stronger SNPs associated with the trait variation were detected for total carotenoids, flour yellowness, and lightness, and δ-tocopherol and α-tocopherol variation.

For both flour's yellowness (*b**) and total carotenoids content (TCC) the strongest SNP associated with both traits was rs131576886 ($-\log_{10}$ (P-value) = 17.648, for *b**, and ($-\log_{10}$ (P-value) = 12.890, for TCC). For *L** the strongest SNP associated with the trait variation was rs131576883 ($-\log_{10}$ (P-value) = 6.384). Both SNPs were not mapped within any gene. Nevertheless, they were located 1,141 bp (rs131576886) and 476 bp (rs131576883) upstream of the GRMZM2G300348 (*y1 - yellow endosperm1*) gene. This gene codes for a phytoene synthase, an enzyme involved in the carotenoid biosynthesis pathway. Considering the local LD decay in the region where those SNPs where located, other nine genes were mapped

within this mega-region (chromosome 6: 78,981 kb to 82,864 kb). Details on their identification as well as functional annotation can be found in Table S8.

For δ-tocopherol, on chromosome 1 (298,814 to 298,815 kb) two SNPs were equally strongly associated with DTLOG variation (rs128990610 and rs128990613; $-\log_{10}$ (P-value) = 5.986). Both SNPs were located within the GRMZM5G876146 (*umc2244*) gene, coding for a pantoate--beta-alanine ligase, an enzyme involved both in the beta-alanine metabolism and in the pantothenate and CoA biosynthesis pathway.

For α-tocopherol, on chromosome 5 (200,420 kb to 200,421 kb) two SNPs were equally strongly associated with ATLOG variation (rs130180529 and rs130180536; $-\log_{10}$ (P-value) = 5.639). Both SNPs were located within the GRMZM2G035213 gene (*vte4 - vitamin E synthesis4*), coding for tocopherol O-methyltransferase, an enzyme involved in vitamin E biosynthesis.

4 Discussion

This work reports the identification of 73 genomic regions associated with the 10 antioxidant compounds-related traits evaluated in wholemeal maize flour. This was achieved through a genome-wide association analysis undertaken using an association panel containing maize inbred lines derived from traditional Portuguese maize populations. This study allowed to identify candidate genes for the majority of the quality associated genomic regions controlling for maize antioxidant compounds-related traits (carotenoids, tocopherols, and phenolic compounds) and flour color. However, also novel regions, with no clear candidates, were identified that were not previously acknowledged using other germplasm collections studies. The association panel showed to be more suitable to study the

genetic architecture of traits either with high heritability values, controlled by a smaller set of genes, and/or traits controlled by large-effect *loci* (e.g., flour yellowness and total carotenoids content).

Especially for δ-tocopherol levels in the inbred lines collection, an environmental influence was observed. This is in line with the genotype-by-environment (G × E) interactions previously described for tocopherol levels in maize kernel (e.g., Chander et al., 2008b).

Several of the SNP-trait associations detected in the present study were located within or near genes known to be involved in the biosynthetic pathway of the compounds under analysis. This observation strengthened and served as a proof-of-concept for the usefulness of the used association panel, though the statistical power to detect the significant associations was clearly constrained by the size of the association panel and by the fast LD decay rate observed in the majority of regions associated with the traits analyzed.

Considering all the regions identified in this work, the genomic region harboring the strongest SNP-traits associations was found on chromosome 6 and was associated with total carotenoids content, flour yellowness, and flour lightness. The strongest SNP-trait associations for total carotenoids content, flour yellowness and flour lightness were not mapped within any gene but were located respectively 1,141 base pairs and 476 base pairs upstream of the GRMZM2G300348 gene (*y1 - yellow endosperm1*), coding for a phytoene synthase (PSY1), an enzyme catalyzing the first committed step of the carotenoids biosynthetic pathway (Buckner et al., 1996). GRMZM2G300348 gene was also identified previously under a QTL controlling for carotenoids levels (Fu et al., 2013; Jittham et al., 2017; Wong et al., 2004) and kernel color (Chandler et al., 2013; Romay et al., 2013).

The high number of strongly significant SNPs both associated with carotenoids levels and flour yellowness found on chromosome 6 near the region harboring the GRMZM2G300348 gene goes in line with the fact that very extensive LD has been previously found around this gene (Palaisa et al., 2004). Extensive LD has been found common in regions that have experienced strong selective sweeps (e.g., Tian et al., 2009). This selection pressure has caused the LD around this locus on yellow maize to span hundreds of kilobase pairs (Yan et al., 2011).

Other genomic regions identified in this work harbored potential candidate genes for which we had no previous information of their involvement with the quality traits analyzed. This was the case for one of the genomic regions on chromosome 1 strongly associated with flour lightness (*L**) (chr1: 76,243 kb to 76,322 kb). Flour lightness is negatively correlated with other two traits measured in this work − flour yellowness (*b**) and total carotenoids content (TCC). Moreover, this region was co-localized with a QTL previously identified for the ratio of β-cryptoxanthin in relation to total carotenoids content (Jittham et al., 2017). The identification of regions containing no obvious candidate genes may result from the use of a different association panel harboring different genetic variability, or simply be due to the rapid rate of LD decay observed in the present panel that hampered the identification of the most obvious candidate.

Plants are the primary source of dietary vitamin E, producing tocopherol and tocotrienol derivatives that collectively constitute vitamin E (DellaPenna & Pogson, 2006). Among these derivatives, α-tocopherol has the highest biological activity for human health (Traber & Atkinson, 2007). Thus, increasing the levels of vitamin E in food crops, particularly of α-tocopherol, is the goal of vitamin E biofortification (review in Jiang et al., 2017). In present work, we

identified a region strongly associated with α-tocopherol. This region was located in chromosome 5 and the strongest associated SNP-α-tocopherol was located within the GRMZM2G035213 gene which codes for γ-tocopherol methyltransferase (VTE4), a well-known enzyme involved in the vitamin E biosynthesis, converting γ-tocopherol to α-tocopherol (Shutu et al., 2012). In the last years, several QTL linkage mapping and associations studies have identified this genomic region and dissected it to the genes level and were able to shown the contribution of this gene in the regulation of tocopherol levels (Chander et al., 2008a; Li et al., 2012; Shutu et al., 2012; Lipka et al., 2013; Diepenbrock et al., 2017).

Also, and in line with the observation that the range of tocopherols is positively correlated with oil content (Nuss & Tanumihardjo, 2010), in the present work, several of the genomic regions identified associated with tocopherols variation were co-localized with QTLs previously identified for oil (fat) or fatty acids content in maize kernel. For instance, the region on chromosome 1 (81,984,649-82,033,180 bp) associated with γ-tocopherol variation co-localized with QTLs for oil and linoleic acid variation in maize kernels (Yang et al., 2010) and also the region on chromosome 1 (298,814,922-298,815,104 bp) strongly associated with δ-tocopherol co-localized with QTLs previously identified for oil variation in maize kernels (Cook et al., 2012).

The Portuguese traditional maize populations were characterized by low yields, besides the low α-tocopherol levels observed in maize flour (Alves et al., Chapter III). Taking this into consideration in the case of a quality-oriented breeding program for maize food using the Portuguese germplasm, one of the possible breeding objectives to pursue would be to increase their limiting tocopherol levels. Additionally, tocopherols present in the seed play

essential physiological roles in the plant as they are involved in guaranteeing seed longevity, preventing lipid peroxidation during germination, and in abiotic stress tolerance (e.g., Chen et al., 2016b; Sattler et al., 2004; Wang et al., 2017). In the present work, two SNPs were strongly associated with α-tocopherol and explained 13.3% of the phenotypic variance observed on chromosome 5. In relation to the average value of the association panel, the effect of the variant allele “AT” led to a decrease of 15.2% - 21.2% in α-tocopherol levels. Additionally, we observed that the allele “AT” that was indeed the most frequent allele in the 29 inbred lines derived entirely from Portuguese traditional maize populations, directing them towards a decrease in levels of α-tocopherol. Therefore, this gene is a promising target for the development of a molecular marker that will aid in the section of lines/populations with higher levels of α-tocopherol.

Noteworthy the mention of one of the SNP significantly associated with *L** variation that was mapped within the GRMZM2G152135 gene. Previous works (Vallabhaneni et al., 2009; Yan et al., 2010; Azmach et al., 2013; Fu et al., 2013; Owens et al., 2014; Suwarno et al., 2015; Jittham et al., 2017) had demonstrated in maize that the gene GRMZM2G152135 (*hyd3* - *hydroxylase3*) located on chromosome 10 and coding for a beta-carotene 3-hydroxylase, an enzyme involved in the carotenoid biosynthesis, underlies a principal quantitative trait locus associated with β-carotene concentration and conversion in maize kernels. Flour lightness was moderately and negatively correlated with total carotenoids content and additionally, these traits also shared several genomic regions associated with their variation on chromosome 6. Therefore, though this region was not identified in this work associated directly with carotenoids variation, it should also be considered as a potential target region driving carotenoid content in this inbred line collection.

Especially for carotenoids (and color yellow/orange), and tocopherols levels in maize kernel, candidate genes that have been consistently identified under QTLs controlling for those compounds in other works such as the *viviparous9* (*vp9*) (Wong et al., 2004; Chandler et al., 2013), *lycopene epsilon cyclase1* (*lycE1*) (Harjes et al., 2008; Chandler et al., 2013; Fu et al., 2013; Owens et al., 2014; Suwarno et al., 2015), *zeaxanthin epoxidase1* (*zep1*) (Chandler et al., 2013; Owens et al., 2014; Suwarno et al., 2015), or the *white cap1* (*wc1*) (Chandler et al., 2013; Suwarno et al., 2015), known for their involvement in the control of kernel carotenoids content and kernel color; as well as the *4-hydroxyphenylpyruvate dioxygenase 1* (*hppd1*) (Chander et al. 2008a), *albino or pale green mutant1* (*apg1, vte3*), and the *homogentisate geranylgeranyl transferase1* (*hggt1*) (Diepenbrock et al., 2017), known for significant association with the variation in kernel tocopherols content, were not detected as associated with that traits variation in the present study. The genotyping platform used on the current work screened several SNPs located within all the aforementioned candidate genes. Nevertheless, no association was detected between those SNPs and the levels of carotenoids, yellowness, and tocopherols in maize flour on the present association panel. As pointed out by Cook et al. (2012) several factors could be responsible for differences in position and quantity of QTLs detected between studies, including variation in allelic frequency, mapping resolution influenced by the magnitude of linkage disequilibrium in a population, marker density, environmental effects, and QTL analysis methods. The relatively small size of the used association panel might have constrained the statistical power to detect significant marker-trait associations in the present study. Also, the rapid rate of LD decay observed in the present study in the SNPs associated with the quality traits evaluated suggests that a

higher marker density would have been beneficial in the detection of other regions putatively linked to maize flour's quality.

The SNPs strongly associated with the traits analyzed and/or the SNPs which allelic variants were found to contribute to larger phenotypic effects should be prioritized as candidate genomic regions for marker development to support selection activities especially for the quality-related traits more difficult to measure/assess. Nevertheless and as already mentioned in Chapter IV, those associations need to be further validated. Future work will concentrate on the validation of the results retrieved in this work by sequencing those regions on contrasting maize populations for the given trait. Since the actual materials used for the manufacturing of the *broa* maize-based bread are the maize populations, these are the ideal independent materials to proceed with the missing validation.

Acknowledgments

MLA performed the DNA isolation, the genotypic and phenotypic data analysis, the association mapping analysis and follow-up analysis, and drafted the manuscript. AB-S and BC performed the quality data acquisition. DG, MP, and CB participated in the establishment and maintenance of the field trials and performed the controlled pollinations. PMM coordinated the field trials and participated in the manuscript revision. CB and MRB coordinated and participated in the acquisition of quality data, and participated in the manuscript revision. MM and FvE participated in the molecular and phenotypic data processing, in the genome-wide association analysis, and participated in the manuscript revision. MCVP designed and coordinated the study and participated in the drafting and revising of the manuscript.

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) and Horizon 2020 Framework Programme under grant agreements No 245058 and No 633571, respectively, and from the Fundação para a Ciência e a Tecnologia, through the research unit GREEN-IT (UID/Multi/04551/2013), and through the research project under the grant agreement no. PTDC/AGR-ALI/099285/2008. MLA and MCVP were supported by Fundação para a Ciência e a Tecnologia (SFRH/BD/77723/2011; FCT Investigator 2015 contract - IF/01337/2014, respectively). The authors would also like to acknowledge the Portuguese Bank of Plant Germplasm for providing the original seeds of the maize inbred lines.

References

AACC International (2012). Approved methods of analysis Method 14-60.01, 11th ed. Total carotenoid content of cereal grains and flours. St.Paul, MN, USA: AACC International.

Adom, K.K., & Liu, R.H. (2002). Antioxidant activity of grains. *Journal of Agricultural and Food Chemistry*, 50(21), 6182-6187.

Azmach, G., Gedil, M., Menkir, A., & Spillane, C. (2013). Marker-trait association analysis of functional gene markers for provitamin A levels across diverse tropical yellow maize inbred lines. *BMC plant biology*, 13(1), 227.

Bento-Silva, A., Patto, M. C. V., & do Rosário Bronze, M. (2017). Relevance, structure and analysis of ferulic acid in maize cell walls. *Food Chemistry*, *In Press*.

Beta, T., & Corke, H. (2004). Effect of ferulic acid and catechin on sorghum and maize starch pasting properties. *Cereal Chemistry*, 81(3), 418-422.

Brites, C., Trigo, M. J., Santos, C., Collar, C., & Rosell, C. M. (2010). Maize-based gluten-free bread: influence of processing parameters on sensory and instrumental quality. *Food and Bioprocess Technology*, 3(5), 707-715.

Buckner, B., San Miguel, P., Janick-Buckner, D., & Bennetzen, J. L. (1996). The *y1* gene of maize codes for phytoene synthase. *Genetics*, 143(1), 479-488.

Carbas, B., Vaz Patto, M. C., Bronze, M. R., Bento da Silva, A., Trigo, M. J., & Brites, C. (2016). Maize flour parameters that are related to the consumer perceived quality of 'broa' specialty bread. *Food Science and Technology*, 36, 259–267.

Cauvain, S. P. (2012). Introducing to bread making. In "Breadmaking: Improving quality". 2nd edition. S. P. Cauvain (Ed.), pp. 1–8. Woodhead: Cambridge, England.

Chander, S., Guo, Y. Q., Yang, X. H., Yan, J. B., Zhang, Y. R., Song, T. M., et al. (2008a). Genetic dissection of tocopherol content and composition in maize grain using quantitative trait loci analysis and the candidate gene approach. *Molecular Breeding*, 22(3), 353.

Chander, S., Meng, Y., Zhang, Y., Yan, J., & Li, J. (2008b). Comparison of nutritional traits variability in selected eighty-seven inbreds from Chinese maize (*Zea mays* L.) germplasm. *Journal of Agricultural and Food Chemistry*, 56(15), 6506-6511.

Chandler, K., Lipka, A. E., Owens, B. F., Li, H., Buckler, E. S., Rocheford, T., et al. (2013). Genetic analysis of visually scored orange kernel color in maize. *Crop Science*, 53(1), 189-200.

Chen, D., Li, Y., Fang, T., Shi, X., & Chen, X. (2016b). Specific roles of tocopherols and tocotrienols in seed longevity and germination tolerance to abiotic stress in transgenic rice. *Plant Science*, 244, 31-39.

Chen, M., Rao, R. S. P., Zhang, Y., Zhong, C., & Thelen, J. J. (2016a). Metabolite variation in hybrid corn grain from a large-scale multisite study. *The Crop Journal*, 4(3), 177-187.

Cook, J. P., McMullen, M. D., Holland, J. B., Tian, F., Bradbury, P., Ross-Ibarra, J., et al. (2012). Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. *Plant Physiology*, 158(2), 824-834.

DellaPenna, D., & Pogson, B. J. (2006). Vitamin synthesis in plants: tocopherols and carotenoids. *Annual Review of Plant Biology*, 57, 711-738.

Diepenbrock, C. H., Kandianis, C. B., Lipka, A. E., Magallanes-Lundback, M., Vaillancourt, B., Góngora-Castillo, E., et al. (2017). Novel Loci Underlie Natural Variation in Vitamin E Levels in Maize Grain. *The Plant Cell* Online, tpc-00475.

Food and Agriculture Organization of the United Nations (FAO). (2012). FAOSTAT. Rome, Italy: FAO.

Fernandes, T. J., Oliveira, M. B. P., & Mafra, I. (2013). Tracing transgenic maize as affected by breadmaking process and raw material for the production of a traditional maize bread, broa. *Food Chemistry* 138(1), 687-692.

Fu, Z., Chai, Y., Zhou, Y., Yang, X., Warburton, M. L., Xu, S., et al. (2013). Natural variation in the sequence of *PSY1* and frequency of favorable polymorphisms among tropical and temperate maize germplasm. *Theoretical and Applied Genetics*, 126(4), 923-935.

Ganal, M. W., Durstewitz, G., Polley, A., Bérard, A., Buckler, E. S., Charcosset, A., et al. (2011). A large maize (*Zea mays* L.) SNP genotyping array: development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. *PloS one*, 6(12), e28334.

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., et al. (2011). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*, 40(D1), D1178-D1186.

Harjes, C. E., Rocheford, T. R., Bai, L., Brutnell, T. P., Kandianis, C. B., Sowinski, S. G., et al. (2008). Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science*, 319(5861), 330-333.

ISO 9936 (2006). Animal and vegetable fats and oils – Determination of tocopherol and tocotrienol contents by high-performance liquid chromatography.

Jiang, L., Wang, W., Lian, T., & Zhang, C. (2017). Manipulation of Metabolic Pathways to Develop Vitamin-Enriched Crops for Human Health. *Frontiers in Plant Science*, 8, 937.

Jittham, O., Fu, X., Xu, J., Chander, S., Li, J., & Yang, X. (2017). Genetic dissection of carotenoids in maize kernels using high-density single nucleotide polymorphism markers in a recombinant inbred line population. *The Crop Journal*, 5(1), 63-72.

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1), 27-30.

Klepacka, J., & Fornal, Ł. (2006). Ferulic acid and its position among the phenolic compounds of wheat. *Critical Reviews in Food Science and Nutrition*, 46(8), 639-647.

Ktenioudaki, A., Alvarez-Jubete, L., & Gallagher, E. (2015). A review of the process-induced changes in the phytochemical content of cereal grains: the breadmaking process. *Critical Reviews in Food Science and Nutrition* 55(5), 611-619.

Lawrence, C. J., Harper, L. C., Schaeffer, M. L., Sen, T. Z., Seigfried, T. E., & Campbell, D. A. (2008). MaizeGDB: The Maize Model Organism Database for Basic, Translational, and Applied Research. *International Journal of Plant Genomics*, 2008:496957.

Li, Q., Yang, X., Xu, S., Cai, Y., Zhang, D., Han, Y., et al. (2012). Genome-wide association studies identified three independent polymorphisms associated with α-tocopherol content in maize kernels. *PLoS One*, 7(5), e36807.

Listgarten, J., Lippert, C., Kadie, C. M., Davidson, R. I., Eskin, E., & Heckerman, D. (2012). Improved linear mixed models for genome-wide association studies. *Nature Methods*, 9(6), 525-526.

Lipka, A. E., Gore, M. A., Magallanes-Lundback, M., Mesberg, A., Lin, H., Tiede, T., et al. (2013). Genome-wide association study and pathway-level analysis of tocochromanol levels in maize grain. G3: Genes, Genomes, Genetics, 3(8), 1287-1299.

Lopez-Martinez, L.X., Oliart-Ros, R.M., Valerio-Alfaro, G., Lee, C.H., Parkin, K.L., & Garcia, H.S. (2009). Antioxidant activity, phenolic compounds and anthocyanins content of eighteen strains of Mexican maize. *Food Science and Technology* 42(6), 1187-1192.

Malosetti, M., van der Linden, C. G., Vosman, B., & van Eeuwijk, F. A. (2007). A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. *Genetics*, 175(2), 879-889.

Matos, M.E., & Rosell, C.M. (2012). Relationship between instrumental parameters and sensory characteristics in gluten-free breads. *European Food Research and Technology*, 235(1), 107-117.

Moose, S. P., & Mumm, R. H. (2008). Molecular plant breeding as the foundation for 21st century crop improvement. *Plant Physiology*, 147(3), 969-977.

Murray, D., Payne, R, & Zhang, Z. (2014) Breeding View, a Visual Tool for Running Analytical Pipelines: User Guide. VSN International Ltd.

Muzhingi, T., Palacios-Rojas, N., Miranda, A., Cabrera, M. L., Yeum, K. J., & Tang, G. (2017). Genetic variation of carotenoids, vitamin E and phenolic compounds in Provitamin A biofortified maize. *Journal of the Science of Food and Agriculture*, 97(3), 793-801.

Naczk, M., & Shahidi, F. (2006). Phenolics in cereals, fruits and vegetables: Occurrence, extraction and analysis. *Journal of Pharmaceutical and Biomedical Analysis*, 41(5), 1523-1542.

Nuss, E. T., & Tanumihardjo, S. A. (2010). Maize: a paramount staple crop in the context of global nutrition. *Comprehensive Reviews in Food Science and Food Safety*, 9(4), 417-436.

Owens, B. F., Lipka, A. E., Magallanes-Lundback, M., Tiede, T., Diepenbrock, C. H., Kandianis, C. B., et al. (2014). A foundation for provitamin A biofortification of maize: genome-wide association and genomic prediction models of carotenoid levels. *Genetics*, 198(4), 1699-1716.

Pei, K., Ou, J., Huang, J., & Ou, S. (2016). *p*-Coumaric acid and its conjugates: dietary sources, pharmacokinetic properties and biological activities. *Journal of the Science of Food and Agriculture*, 96(9), 2952-2962.

Palaisa, K., Morgante, M., Tingey, S., & Rafalski, A. (2004). Long-range patterns of diversity and linkage disequilibrium surrounding the maize *Y1* gene are indicative of an asymmetric selective sweep. *Proceedings of the National Academy of Sciences of the United States of America*, 101(26), 9885-9890.

Payne R., Murray D., & Harding S. (2015). GenStat for Windows (18th Edition) Introduction. *VSN International: Hemel Hempstead*, UK.

Ranum, P., Peña-Rosas, J. P., & Garcia-Casal, M. N. (2014). Global maize production, utilization, and consumption. *Annals of the New York Academy of Sciences*, 1312(1), 105-112.

Rincent, R., Moreau, L., Monod, H., Kuhn, E., Melchinger, A. E., Malvar, R. A., et al. (2014). Recovering power in association mapping panels with variable levels of linkage disequilibrium. *Genetics*, 197(1), 375-387.

Rodríguez, V. M., Soengas, P., Landa, A., Ordás, A., & Revilla, P. (2013). Effects of selection for color intensity on antioxidant capacity in maize (Zea mays L.). *Euphytica* 193(3), 339-345.

Romay, M. C., Millard, M. J., Glaubitz, J. C., Peiffer, J. A., Swarts, K. L., Casstevens, T. M., et al. (2013). Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biology*, 14(6), R55.

Salinas-Moreno, Y., García-Salinas, C., Ramírez-Díaz, J. L., & Alemán-de la Torre, I. (2017). Phenolic compounds in maize grains and its nixtamalized products. In *Phenolic Compounds-Natural Sources, Importance and Applications*. Rijeka, Croatia: InTech.

Santiago, R., Malvar, R. A., Barros-Rios, J., Samayoa, L. F., & Butrón, A. (2016). Hydroxycinnamate synthesis and association with Mediterranean corn borer resistance. *Journal of Agricultural and Food Chemistry*, 64(3), 539-551.

Sattler, S.E., Gilliland, L.U., Magallanes-Lundback, M., Pollard, M., & DellaPenna, D. (2004). Vitamin E is essential for seed longevity and for preventing lipid peroxidation during germination. *The Plant Cell* 16(6), 1419-1432.

Schnable, P. S., Ware, D., Fulton, R. S., Stein, J. C., Wei, F., Pasternak, S., et al. (2009). The B73 maize genome: complexity, diversity, and dynamics. *Science*, 326(5956), 1112-1115.

Shahidi, F. (Ed.). (1997). Natural antioxidants: chemistry, health effects, and applications. The American Oil Chemists Society Press, Champaign, Illinois, USA.

Shutu, X., Dalong, Z., Ye, C., Yi, Z., Shah, T., Ali, F., et al. (2012). Dissecting tocopherols content in maize (*Zea mays* L.), using two segregating populations and high-density single nucleotide polymorphism markers. *BMC Plant Biology*, 12(1), 201.

Silva, S., Bronze, M.R., Figueira, M.E., Siwy, J., Mischak, H., Combet, E., et al. (2015). Impact of a 6-wk olive oil supplementation in healthy adults on urinary proteomic biomarkers of coronary artery disease, chronic kidney

disease, and diabetes (types 1 and 2): A randomized, parallel, controlled, double-blind study. *The American Journal of Clinical Nutrition*, 101(1), 44-54.

Singleton, V.L., Orthofer, R., & Lamuela-Raventos, R.M. (1999). Analysis of total phenols and other oxidant substrates and antioxidants by means of Folin-Ciocalteu reagent. *Methods in Enzymology*, 299, 157-178.

Siriamornpun, S., Tangkhawanit, E., & N. Kaewseejan. (2016). Reducing retrogradation and lipid oxidation of normal and glutinous rice flours by adding mango peel powder. *Food Chemistry*, 201, 160-167.

Siró, I., Kápolna, E., Kápolna, B., & Lugasi, A. (2008). Functional food. Product development, marketing and consumer acceptance—A review. *Appetite* 51(3), 456-467.

Suwarno, W. B., Pixley, K. V., Palacios-Rojas, N., Kaeppler, S. M., & Babu, R. (2015). Genome-wide association analysis reveals new targets for carotenoid biofortification in maize. *Theoretical and Applied Genetics*, 128(5), 851-864.

Tenaillon, M.I., & Charcosset, A. (2011). A European perspective on maize history. *Comptes Rendus Biologies*, 334(3), 221-228.

The IBP Breeding Management System Version 3.0.9 (December 2015) The Integrated Breeding Platform. Available at <https://www.integratedbreeding.net/breeding-management-system>

Tian, F., Stevens, N. M., & Buckler, E. S. (2009). Tracking footprints of maize domestication and evidence for a massive selective sweep on chromosome 10. *Proceedings of the National Academy of Sciences*, 106(Supplement 1), 9979-9986.

Traber, M. G., & Atkinson, J. (2007). Vitamin E, antioxidant and nothing more. Free *Radical Biology and Medicine*, 43(1), 4-15.

Treutter, D. (2010). Managing phenol contents in crop plants by phytochemical farming and breeding—visions and constraints. *International Journal of Molecular Sciences*, 11(3), 807-857.

Vallabhaneni, R., Gallagher, C. E., Licciardello, N., Cuttriss, A. J., Quinlan, R. F., & Wurtzel, E. T. (2009). Metabolite sorting of a germplasm collection reveals the *hydroxylase3* locus as a new target for maize provitamin A biofortification. *Plant Physiology*, 151(3), 1635-1645.

Vaz Patto, M.C., Alves, M.L., N.F., Almeida, C.S., Mendes, P., & Satovic, Z. (2009). Is the bread making technological ability of Portuguese traditional maize landraces associated with their genetic diversity? *Maydica*, 54(2/3), 297-311.

Vaz Patto, M. C., Moreira, P. M., Carvalho, V., & Pego, S. (2007). Collecting maize (*Zea mays* L. convar. *mays*) with potential technological ability for bread making in Portugal. *Genetic Resources and Crop Evolution*, 54(7), 1555-1563.

Vaz Patto, M. C., Satovic, Z., Pego, S., & Fevereiro, P. (2004). Assessing the genetic diversity of Portuguese maize germplasm using microsatellite markers. *Euphytica*, 137(1), 63-72.

Wang, D., Wang, Y., Long, W., Niu, M., Zhao, Z., Teng, X., et al. (2017). SGD1, a key enzyme in tocopherol biosynthesis, is essential for plant development and cold tolerance in rice*. Plant Science*, 260, 90-100.

Wurtzel, E. T., Cuttriss, A., & Vallabhaneni, R. (2012). Maize provitamin A carotenoids, current resources, and future metabolic engineering challenges. *Frontiers in Plant Science*, 3.

Wen, W., Brotman, Y., Willmitzer, L., Yan, J., & Fernie, A. R. (2016a). Broadening our portfolio in the genetic improvement of maize chemical composition. *Trends in Genetics*, 32(8), 459-469.

Wen, W., Li, D., Li, X., Gao, Y., Li, W., Li, H., et al. (2014). Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nature communications*, 5.

Wen, W., Liu, H., Zhou, Y., Jin, M., Yang, N., Li, D., et al. (2016b). Combining quantitative genetics approaches with regulatory network analysis to dissect the complex metabolism of the maize kernel. *Plant Physiology*, 170(1), 136-146.

Wong, J. C., Lambert, R. J., Wurtzel, E. T., & Rocheford, T. R. (2004). QTL and candidate genes phytoene synthase and ζ-carotene desaturase associated with the accumulation of carotenoids in maize. *Theoretical and Applied Genetics*, 108(2), 349-359.

Yan, J., Kandianis, C. B., Harjes, C. E., Bai, L., Kim, E. H., Yang, X., et al. (2010). Rare genetic variation at Zea mays crtRB1 increases [beta]-carotene in maize grain. *Nature Genetics*, 42(4), 322-327.

Yan, J., Warburton, M., & Crouch, J. (2011). Association mapping for enhancing maize (*Zea mays* L.) genetic improvement. *Crop Science*, 51(2), 433-449.

Yang, X., Guo, Y., Yan, J., Zhang, J., Song, T., Rocheford, T., et al. (2010). Major and minor QTL and epistasis contribute to fatty acid compositions and oil concentration in high-oil maize. *Theoretical and Applied Genetics*, 120(3), 665-678.

Zhai, S., Xia, X., & He, Z. (2016). Carotenoids in staple cereals: metabolism, regulation, and genetic manipulation. *Frontiers in Plant Science*, 7.

Zhu, F., Cai, Y.Z., Sun, M., & Corke, H. (2009). Effect of phytochemical extracts on the pasting, thermal, and gelling properties of wheat starch. *Food Chemistry*, 112(4), 919-923.

Žilić, S., Serpen, A., Akıllıoğlu, G., Gökmen, V., & Vančetović, J. (2012). Phenolic compounds, carotenoids, anthocyanins, and antioxidant capacity of colored maize (*Zea mays* L.) kernels. *Journal of Agricultural and Food Chemistry*, 60(5), 1224-1231.

Supplementary material

Tables

Table S1. Maize inbred lines with available quality data, known pedigree, kernel color, and endosperm type. – Table available online through the link <https://figshare.com/s/051f42ca548266f113e2>

Table S2. Pearson correlation coefficients among quality traits measured in wholemeal flour of 132 maize inbred lines. The phenotypic correlations were calculated independently for each growing season evaluated (2011 and 2012). The values above the diagonal correspond to the phenotypic correlations among quality traits measured in the first growing season (2011); values below the diagonal correspond to the phenotypic correlations among quality traits measured in the second growing season (2012). In bold are highlighted the strong phenotypic correlations (|r| > 0.8).

		*L**	*a**	*b**	TCC	ATLOG	DTLOG	GT	PHSLOG	PHHLOG	HYLOG
*L**	r	-	-0.36	-0.51	-0.53	0.08	-0.01	-0.16	-0.21	-0.27	0.08
	P-value	-	***	***	***	ns	ns	ns	ns	*	ns
*a**	r	-0.48	-	-0.16	-0.12	-0.16	-0.04	-0.09	0.00	0.05	-0.05
	P-value	***	-	ns	ns	ns	ns	ns	ns	ns	ns
*b**	r	-0.60	-0.18	-	**0.93**	0.06	-0.04	0.11	0.18	0.07	-0.03
	P-value	***	ns	-	***	ns	ns	ns	ns	ns	ns
TCC	r	-0.60	-0.12	**0.96**	-	0.06	-0.02	0.13	0.22	0.15	-0.04
	P-value	***	ns	***	-	ns	ns	ns	*	ns	ns
ATLOG	r	-0.05	-0.20	0.36	0.39	-	-0.12	0.01	0.02	-0.02	0.11
	P-value	ns	ns	***	***	-	ns	ns	ns	ns	ns
DTLOG	r	0.32	-0.18	-0.06	-0.09	0.09	-	0.72	-0.01	0.13	0.16
	P-value	**	ns	ns	ns	ns	-	***	ns	ns	ns
GT	r	0.19	-0.19	0.10	0.08	0.17	0.54	-	0.05	0.11	0.07
	P-value	ns	ns	ns	ns	ns	***	-	ns	ns	ns
PHSLOG	r	-0.07	-0.06	0.07	0.12	0.07	0.02	0.20	-	0.64	0.21
	P-value	ns	ns	ns	ns	ns	ns	ns	-	***	ns
PHHLOG	r	-0.17	0.01	0.00	0.10	0.10	-0.06	0.09	**0.78**	-	0.49
	P-value	ns	ns	ns	ns	ns	ns	ns	***	-	***
HYLOG	r	0.15	0.01	-0.21	-0.17	0.00	0.11	0.17	0.55	0.60	-
	P-value	ns	ns	ns	ns	ns	ns	ns	***	***	-

*r corresponds to the Person correlation coefficient; P-value corresponds to the significance level of correlations indicated as: ns - non-significant; * - significant at P < 0.05; ** - significant at P < 0.01; *** - significant at P < 0.001.*

Quality traits' abbreviations: L – flour's lightness; a* – flour's red/green index, positive values indicate that samples tend toward the red part of the color spectra; b* – flour's yellow/blue index, positive values indicates that samples tend toward the yellow part of the color spectra; TCC – total carotenoids content; ATLOG – log10-transformed α-tocopherol content; DTLOG – log10-transformed δ-tocopherol content; GT – γ-tocopherol content; PHSLOG – log10-transformed total phenolic compounds, assessed by Folin–Ciocalteu assay; PHHLOG – log10-transformed total phenolic compounds, assessed by HPLC; HYLOG – log10-transformed total hydroxycinnamic acids, assessed by HPLC*

Table S3. Estimated genetic correlations among quality traits measured in wholemeal flour of a collection of 132 maize inbred lines evaluated during two growing seasons. The genetic correlations were calculated independently for each growing season (2011 and 2012). Values above the diagonal correspond to the genetic correlations among quality traits measured in the first growing season (2011); values below the diagonal correspond to the genetic correlations among quality traits measured in the second growing season (2012). In bold are highlighted the strong genetic correlations (|r| > 0.8).

	*L**	*a**	*b**	TCC	ATLOG	DTLOG	GT	PHSLOG	PHHLOG	HYLOG
*L**	-	–0.377	–0.480	–0.465	–0.060	0.055	–0.034	–0.175	–0.218	0.037
*a**	–0.518	-	–0.181	–0.114	–0.010	–0.044	–0.097	–0.007	0.066	–0.020
*b**	–0.527	–0.205	-	**0.898**	0.119	–0.051	0.137	0.131	0.038	–0.043
TCC	–0.553	–0.102	**0.894**	-	0.162	–0.060	0.161	0.153	0.164	–0.019
ATLOG	–0.075	–0.124	0.309	0.362	-	–0.094	0.050	0.084	0.083	0.151
DTLOG	0.261	–0.177	–0.017	–0.014	0.106	-	0.698	0.004	0.055	0.132
GT	0.166	–0.181	0.143	0.147	0.166	0.551	-	0.093	0.070	0.079
PHSLOG	–0.031	0.003	–0.046	–0.032	0.115	0.055	0.153	-	0.707	0.398
PHHLOG	–0.118	–0.021	–0.010	0.029	0.070	–0.008	0.115	0.697	-	0.542
HYLOG	0.206	–0.086	–0.176	–0.224	–0.023	0.189	0.145	0.464	0.578	-

Quality traits' abbreviations: L* – *flour's lightness;* a* – *flour's red/green index, positive values indicate that samples tend toward the red part of the color spectra;* b* – *flour's yellow/blue index, positive values indicates that samples tend toward the yellow part of the color spectra; TCC – total carotenoids content; ATLOG – log_{10}-transformed α-tocopherol content; DTLOG – log_{10}-transformed δ-tocopherol content; GT – γ-tocopherol content; PHSLOG – log_{10}-transformed total phenolic compounds, assessed by* Folin–Ciocalteu *assay; PHHLOG – log_{10}-transformed total phenolic compounds, assessed by HPLC; HYLOG – log_{10}-transformed total hydroxycinnamic acids, assessed by HPLC.*

Table S4. Phenotypic values (range, and mean ± standard deviation) for quality traits in wholemeal flour of a collection of 132 maize inbred lines evaluated in two growing seasons (2011 and 2012).

Trait	Range (minimum - maximum)		Mean ± standard deviation	
	2011	2012	2011	2012
*L**, in CIE tristimulus color spectra units	78.72 - 87.69	78.65 - 87.26	83.41 ± 1.74	83.62 ± 1.66
*a**, in CIE tristimulus color spectra units	–1.36 - 2.14	–1.28 - 2.02	0.22 ± 0.68	0.35 ± 0.73
*b**, in CIE tristimulus color spectra units	10.44 - 39.87	10.61 - 38.19	25.95 ± 8.30	24.14 ± 8.63
TCC, in µg of lutein equivalent/gr of sample	1.89 - 102.01	1.06 - 74.02	39.69 ± 26.51	31.80 ± 21.85
ATLOG[1], in µg/g fat	0.10 - 2.51	–0.77 - 2.41	1.48 ± 0.51	1.06 ± 0.69
DTLOG[1], in µg/g fat	–0.37 - 1.50	–0.18 - 1.69	0.61 ± 0.41	0.86 ± 0.38
GT, in µg/g fat	4.29 - 642.93	9.78 - 533.06	242.39 ± 153.25	192.08 ± 119.52
PHSLOG[1] [2], in mg GAE / 100 g DW	1.65 - 2.13	1.73 - 2.15	1.89 ± 0.10	1.96 ± 0.09
PHHLOG[1] [2], in mg GAE / 100 g DW	1.71 - 2.25	1.71 - 2.22	1.99 ± 0.11	1.96 ± 0.11
HYLOG[1] [3], in mg FAE / 100 g DW	–0.47 - 1.38	0.04 - 1.63	0.45 ± 0.40	0.78 ± 0.33

[1] *Log_{10} transformed values*

[2] *mg GAE / 100 g DW – milligrams of gallic acid equivalents per 100 grams of dry weight*

[3] *mg FAE / 100 g DW - milligrams ferulic acid equivalents per 100 grams of dry weight*

Quality traits' abbreviations: L* *– flour's lightness;* a* *– flour's red/green index;* b* *– flour's yellow/blue index; TCC – total carotenoids content; ATLOG – α-tocopherol content; DTLOG – δ-tocopherol content; GT – γ-tocopherol content; PHSLOG – total phenolic compounds, assessed by Folin–Ciocalteu assay; PHHLOG – total phenolic compounds, assessed by HPLC; HYLOG – total hydroxycinnamic acids*

Table S5. Observed inflation factors for the models tested in genome-wide association (GWAS) analysis. Inflation factor for the adaptive kinship model corresponds to the average value across chromosomes.

Trait		Naive	Eigen	Adaptive Kinship [1]
*L**		1.294	1.156	1.114
*a**		1.245	1.092	1.068
*b**		1.859	1.321	1.078
TCC		1.804	1.324	1.095
ATLOG		1.357	1.130	1.027
	2011 growing season	1.082	1.039	1.012
DTLOG[2]	2012 growing season	1.130	1.076	1.012
	across growing seasons	1.048	1.048	1.020
GT		1.302	1.050	1.040
PHSLOG		1.015	1.015	1.011
PHHLOG		1.174	1.126	1.026
HYLOG		1.105	1.059	1.002

[1] *Calculated according to Listgarten et al., 2012; Rincent et al., 2014* | [2] *For* DTLOG, *genome-wide association analysis was performed using the phenotypic adjusted means across growing seasons, and individually for each growing season.*

Quality traits' abbreviations: L* - *flour's lightness;* a* - *flour's red/green index;* b* - *flour's yellow/blue index; TCC - total carotenoids content; ATLOG - α-tocopherol content; DTLOG - δ-tocopherol content; GT - γ-tocopherol content; PHSLOG - total phenolic compounds by* Folin–Ciocalteu *assay; PHHLOG - total phenolic compounds by HPLC; HYLOG - total hydroxycinnamic acids*

Table S6. Significant SNP-trait associations using -log10 (P-value) = 4, as the threshold from a genome-wide association study for 10 health-related quality traits using a collection of 132 maize inbred lines evaluated across two growing seasons. – Table available through the link <https://figshare.com/s/6f7f7dcd7b689a145603>

Table S7. Percentage of associated SNP variants with an effect of decrease and increase of the traits value, and maximum and minimum phenotypic variance explained by the SNPs associated with 10 health-related quality traits.

% SNPs $_{DECREASE}$ – Percentage of associated SNPs for which the effect of the rare allele resulted in a decrease of the trait value when compared to the most frequent allele; %SNPs $_{INCREASE}$ - Percentage of associated SNPs for which the effect of the rare allele results in an increase of the trait value when compared to the most frequent allele; V_{MIN} – smallest proportion of variance explained by a significant SNP; V_{MAX} – largest proportion of variance explained by a significant SNP; SNP_{EFFECT} – Effect of variant allele explaining the largest proportion of the trait variance.

Trait		%SNPs DECREASE	%SNPs INCREASE	V_{MIN} (SNP ID; location)	V_{MAX} (SNP ID; location)	SNP_{EFFECT}
*L**, in CIE color system units		60.00	40.00	0.072 (rs130510981; Chr7: 35,175,353 bp[4])	0.133 (rs131576883; chr6: 82,180,010 bp)	(+)1.017 CIE units
*a**, in CIE color system units		0.00	100.00	0.111 (rs131186816; Chr5: 216,492,480 bp)	0.165 (rs129060682; Chr2: 47,072,374 bp)	(+)0.304 CIE units
*b**, in CIE color system units		46.15	53.85	0.027 (rs131100735; Chr9: 102,277,067 bp)	0.207 (rs131576886; Chr6: 82,179,345 bp)	(−)6.806 CIE units
TCC, in µg of lutein equivalents/g of sample		38.46	61.54	0.038 (rs131620719; Chr7: 47,952,859 bp)	0.179 (rs131576886; Chr6 82,179,345 bp)	(−)17.666 µg/g sample
ATLOG[1], in µg/g fat		42.86	57.14	0.096 (rs128647552; Chr10: 149,506,640 bp)	0.133 (rs130180529 and rs130180536; Chr5:200,420,347-200,421,939 bp)	(−)0.255 µg/g fat
DTLOG[1], in µg/g fat	2011	100.00	0.00	0.103 (rs131462576; Chr4: 71,242,320 bp)	0.128 (rs131335950; Chr2: 41,312,594 bp)	(−)0.172 µg/g fat
	2012	85.71	14.29	0.114 (rs131553173; Chr6: 22,002,210 bp)	0.183 (rs131568350; Chr6: 60,174,784 bp)	(−)0.224 µg/g fat
	across growing seasons	88.89	11.11	0.120 (rs131510695; Chr5: 33,437,545 bp)	0.197 (rs128990610 and rs128990613; Chr1: 298,814,922-298,815,104 bp)	(−)0.272 µg/g fat

Continuation Table S7

Trait	%SNPs DECREASE	%SNPs INCREASE	V_{MIN} (SNP ID; location)	V_{MAX} (SNP ID; location)	SNP_{EFFECT}
GT, in µg/g fat	20.00	80.00	0.072 (rs131280450; Chr1: 101,025,418 bp)	0.095 (rs131269931; Chr1: 82,014,738 bp)	(+)56.45 µg/g fat
PHSLOG[1 2], in mg GAE/100 g DW	100.00	0.00	0.080 (rs132575077; Chr9: 154,122,282 bp)	0.089 (rs131185942; Chr2: 3,067,636 bp)	(−)0.026 mg GAE/100 g DW
PHHLOG[1 2], in mg GAE/100 g DW	0.00	100.00	0.069 (rs130095308; Chr5: 126,063,869 bp)	0.069 (rs130095308; Chr5: 126,063,869 bp)	(+)0.033 mg GAE/100 g DW
HYLOG[1 3], in mg FAE/100 g DW	0.00	100.00	0.091 (rs131186110; Chr1: 195,922,510 bp)	0.096 (rs128811826; Chr1: 176,867,145 bp)	(+)0.127 mg FAE/100 g DW

[1] Log_{10} transformed values | [2] mg GAE / 100 g DW – milligrams of gallic acid equivalents per 100 grams of dry weight | [3] mg FAE / 100 g DW - milligrams ferulic acid equivalents per 100 grams of dry weight | [4] bp – base pairs

Quality traits' abbreviations: L* - *flour's lightness;* a* - *flour's red/green index;* b* - *flour's yellow/blue index; TCC - total carotenoids content; ATLOG - α-tocopherol content; DTLOG - δ-tocopherol content; GT - γ-tocopherol content; PHSLOG - total phenolic compounds by* Folin–Ciocalteu *assay; PHHLOG - total phenolic compounds by HPLC; HYLOG - total hydroxycinnamic acids*

Table S8. Candidate genes mapped within the genomic regions associated with 10 health-related quality traits. – Table available online through the link <https://figshare.com/s/c63a960416b0fc0115f1>

Figures

Figure S1. Manhattan plots showing the genome-wide association results for each of the 10 health-related quality traits obtained using a collection of 132 maize inbred lines evaluated across two growing seasons, and for individual growing seasons in the case of the genome-wide association results for δ-tocopherol. – Figure available online through the link <https://figshare.com/s/110da3a07eda01a98529>

Chapter VI

General discussion

General discussion

Quality is receiving increasing relevance on plant breeding efforts to develop healthier and more nutritious crops (Hefferon, 2015). Maize is one of the main crops used for human consumption and, due to this, in high demand for food purposes (Nuss & Tanumihardjo, 2010; Ranum et al., 2014). Consumers worldwide are increasingly concerned with food quality. Breeding for improved plant quality is, however, a complex task (Jiang et al., 2017; Munck, 2009; Wen et al., 2016) and, therefore, the development of tools that will allow a more efficient and effective selection for better quality products is of great importance nowadays.

In Portugal, a unique germplasm has been developed through centuries of adaptation to local environment and food uses, in particular, for ethnic maize leavened *broa* bread production (reviewed in Vaz Patto et al., 2013). Several parameters related to kernel composition, flour pasting behavior and flour particle size have been previously identified as crucial for *broa* quality (Brites et al., 2010; Carbas et al., 2016). Because of their use for human consumption (*broa* bread), these maize landraces are in part maintained, and not yet totally replaced by commercial hybrids (Vaz Patto et al., 2007). Nonetheless, the underuse of these materials as well as the limited knowledge on their phenotypic and molecular characterization as motivated the work developed under this Ph.D. thesis. In this way, the knowledge generated, through the efforts undertaken with this project, is a genuine attempt to contribute to the conservation promotion and revival the use of the Portuguese traditional maize populations, unveiling their potential for a quality-oriented breeding. In this thesis, molecular markers were used, together with phenotypic (agronomic and quality) data, to evaluate the effect of on-farm

stratified mass selection in two historical maize populations selected under the a long-term participatory breeding program - the VASO program; and to characterize the genetic diversity of Portuguese maize populations still under cultivation (farmers' populations). Additionally, a maize inbred line collection partially derived from Portuguese maize populations was analyzed at the molecular and phenotypic level to perform a whole-genome association study, in order to scrutinize the complex genetic basis and identify genomic regions/candidate genes associated with maize bread quality. These molecular-based tools would be fundamental for future maize breeding given the difficulty to visually select for the majority of the quality-related traits.

The main achievements of this Ph.D. project were:

(1) To provide further evidence for the effectiveness of participatory breeding methodologies on agronomic plant improvement, while maintaining high molecular diversity, in two historical maize open-pollinated populations, *Amiúdo* and *Castro Verde* (Chapter II). Our observations also bring awareness for the need to develop selection tools for characteristics that cannot be visually selected by farmers, in order to trace down or improve these traits. The development of these missing tools was further explored in Chapter IV and V.

(2) Through the integration of both phenotypic and genotypic characterization gathered throughout Chapter III, to generate a valuable tool to support an efficient and effective management of the available genetic resources in future breeding activities.

(3) By employing a genome-wide association approach on a maize inbred collection containing lines partially derived from Portuguese maize populations, to unveil the genetic basis of the quality traits

evaluated in Chapters IV and V. A total of 128 genomic regions were identified associated with the different compositional and pasting behavior quality traits, and the different health-related quality traits evaluated.

The Portuguese traditional landraces have been evolving since maize introduction in the country and can still be found under cultivation at the farmers' fields (Vaz Patto et al., 2007). Generally, landraces are known to be less productive than hybrid varieties (Revilla et al., 2015) so their agronomic improvement (yield) is always an important aspect to be considered in future breeding activities. Moreover, maize landraces are considered to have a broader plasticity to adapt to different environments (Hellin et al., 2014) and given the present climate changes concerns (Wheeler & von Braun, 2013) those materials could represent a valuable asset to breed for unpredictable environments.

Traditional maize populations collected from the Central Portuguese region, known to produce a market renowned maize-based bread (Vaz Patto et al., 2007), are not currently involved in any conventional breeding program, neither on the long-term Portuguese participatory maize breeding program (VASO program) (Vaz Patto et al., 2013). Taking into account their potential to improve maize quality-related aspects, a similar breeding methodology as currently used in the Portuguese participatory breeding program (for example, stratified mass selection) could be applied to these populations. However, an inclusive evaluation of the effect of stratified mass selection methodology on an extended number of maize populations was at the beginning of this thesis still missing. In Chapter II, taking advantage of the material selected through the VASO program, we expanded the knowledge on the effects of stratified mass selection

methodology at the agronomic, quality and molecular levels on the historical Portuguese maize populations.

Also in Chapter II, we compared the agronomic performance of the different populations and their selection cycles in multi-location field trials. Multi-location field trials, comparing the initial populations with the derived selection cycles, showed that this selection methodology led to agronomic improvement in one of the populations. In the literature, some examples showing the potential of stratified mass selection specifically in the context of a participatory maize breeding program can be found described in Mendes-Moreira et al. (2008, 2009) and Smith et al. (2001). Since the two populations analyzed on Chapter II are used for human consumption, we also measured several traits associated with grain quality on the same material harvested from a single field location. This analysis showed that the majority of the quality traits evaluated progressed erratically over time during selection. We assessed as well the evolution of the molecular diversity along the selection process using microsatellite markers. The molecular diversity analysis revealed that the overall genetic diversity in both populations was maintained throughout the selection. One of the reasons for the maintenance of the overall genetic diversity levels of both populations can be due to their effective populations' size, which was above 120 individuals. According to Hoban et al. (2014), changes in genetic diversity levels are most likely identified only when the effective population size is smaller than 100 individuals.

Given their specific uses in food, these landraces can be relevant sources of interesting alleles to improve quality of maize-based food products. Previous studies have reported promising differences in the quality of kernels among the farmers' maize populations collected in a Portuguese region known to produce *broa*

bread (Vaz Patto et al., 2009). Several limitations have been identified in the previous characterizations of those populations (Vaz Patto et al., 2009), such as a reduced set of quality traits accessed and a missing accurate agronomic performance evaluation and these were addressed in Chapter III of the present thesis. The results from Chapter III allowed expanding the current knowledge on Portuguese farmers' maize populations collected from a traditional maize-based bread-national producing region. Namely, by generating a more thorough characterization of their phenotypic (quality and agronomic) and genetic diversity that allowed to better organize future breeding activities and identify sources (populations) of interesting traits to be used on future crosses. These maize populations were evaluated for grain yield and ear weight in nine locations across Portugal. The populations' adaptability and stability were evaluated using additive main effects and multiplication interaction (AMMI) model analysis. Regarding the agronomic performance, farmers' maize populations had low but considerably stable grain yields across the tested environments. Hellin et al. (2014) also mentioned the bigger stability of Mexican maize landraces when compared with hybrids; these populations had a wider plasticity to adapt to different environmental conditions while still maintaining yield. The majority of the farmers' populations analyzed in this thesis were characterized by high levels of protein and fiber, low levels of carotenoids, volatile aldehydes, α- and δ-tocopherols content, and low breakdown viscosity values. An example as of how the phenotypic and molecular information collected can be integrated and applied into a decision-making process to support the establishment of a quality-oriented participatory maize breeding program was presented and discussed in Chapter III. Specifically, one of the breeding objectives to be pursued could focus on increasing the agronomic performance of the

populations and tocopherol levels (α- and δ-tocopherol content) that are limiting on this germplasm. An increase in maize vitamin E levels, as tocopherols, can elevate its nutritional value by enhancing their role as antioxidants (Nuss & Tanumihardjo, 2010). Moreover, tocopherols play an essential physiological role in the plant as vitamin E is involved in guaranteeing seed longevity, preventing lipid peroxidation during germination, and in abiotic stress tolerance (e.g., Chen et al., 2016; Sattler et al., 2004; and Wang et al., 2017).

As previously discussed by others, the majority of maize traits, including kernel quality-related traits, have complex patterns of inheritance, being controlled by multiple genes (Wallace et al., 2013). From the results in Chapter II, it was observed that the majority of the quality traits evaluated progressed erratically over time stressing the importance on the development of quality-related molecular selection tools. Moreover, the influence of environmental conditions in phenotypic data related to quality traits has also been reported (e.g., Ketthaisong et al., 2014; Wilson et al., 2004). In Chapters IV and VI of this work we took advantage of the existing maize inbred line collection from the Portuguese Plant Germplasm Bank and used, for the first time, an original collection of 132 maize inbred lines, partially developed from Portuguese traditional maize populations, to carried a genome-wide association study aiming to identify genomic regions/candidate genes controlling compositional and pasting properties of maize wholemeal flour.

This study allowed to better understand the complex genetic basis of maize kernel main compositional and pasting quality, by identifying candidate genes for the majority of the quality associated genomic regions, or the most promising target regions to develop molecular tools to increase efficacy and efficiency of quality selection within maize breeding programs. Important to mention is that the size

of the collection of maize inbred lines used in this work most likely affected the power to detect significant marker-trait associations, and the subsequent identification of genomic regions controlling the analyzed traits association. As reported in Yang et al. (2010), using simulation studies, a collection of 155 diverse maize lines for association mapping was suitable to study mainly traits controlled by major QTLs and the collection size should be extended for further investigating the genetic basis of other traits controlled by genes with moderate or even minor effects. Nevertheless, with this approach, a total of 128 genomic regions associated with the quality traits, evaluated in Chapters III and IV, were identified: 57 genomic regions associated with the 11 different compositional and pasting behavior quality traits evaluated, and 73 genomic regions associated with the 10 different health-related quality traits evaluated.

In our genome-wide association analysis, the strongest marker-trait associations detected were associated with total carotenoids content, flour yellowness, and lightness variation. The strongest SNP-trait associations for these three traits were located upstream of the GRMZM2G300348 gene (*y1* - *yellow endosperm1*), coding for phytoene synthase (PSY1), an enzyme catalyzing the first committed step of the carotenoids biosynthetic pathway (Buckner et al., 1996). GRMZM2G300348 gene was also identified previously under a QTL controlling for carotenoids levels (Fu et al., 2013; Jittham et al., 2017; Wong et al., 2004) and kernel color (Chandler et al., 2013; Romay et al., 2013). Several other regions controlling multiple traits were also detected in the present study with the subsequent identification of potential candidate genes. As an example, for breakdown viscosity and peak viscosity, two viscosity parameters that reflect the starch capacity to absorb water and swell, the strongest common associated region was located near the *dull endosperm 1* gene, which encodes a

starch synthase and is determinant on the starch endosperm structure in maize (Gao et al., 1998; Wu et al., 2015). Other *a priori* candidate gene, the GRMZM2G035213 gene (*vte4*), which codes for γ-tocopherol methyltransferase (VTE4), a well-known enzyme involved in the vitamin E biosynthesis (Shutu et al., 2012), was found under a genomic region associated with α-tocopherol variation. Quantitative trait loci (QTLs) contributing for tocopherol levels have been consistently identified by others in the region where this gene is located (e.g., Diepenbrock et al., 2017; Lipka et al., 2013; Shutu et al., 2012).

Concerning future work related to the detected SNP-trait associations, the regions detected to be associated with the several quality traits analyzed in this work will need to be validated using an independent genetic background before they can become applicable in Marker-Assisted Selection (MAS). Priority will be given to the regions where the strongest association, higher SNP effect size, and where QTLs for the same or related traits were detected on previous works. Moreover, in the case of SNPs located near or within intra-genic regions, the putative functional effect of each associated SNP, using *in silico* prediction based on the maize B73 reference genome, will also serve as criteria to further investigate the possibility of those SNPs being directly linked to the trait, as advised for instance in McLaren et al. (2016).

Currently, a genome-wide association approach is also being undertaken to detect genomic regions involved in the variation of volatile compounds content in maize flour. Specifically, the identification of genomic regions controlling volatile aldehydes content is one of the research topics being presently addressed. The majority of farmers' populations were characterized by low levels of volatile aldehydes (Alves et al., Chapter III). It is widely known that the aroma

strongly influences consumer preference and acceptance of baked goods. Aldehydes have been identified as the main volatile compounds that contribute to the aroma in cereals (Klensporf & Jelén, 2005), and aroma volatiles such as aldehydes resulting from the polyunsaturated fatty acids' oxidation can contribute to the development of off-flavors and rancidity (Gwirtz & Garcia-Casal, 2014). Previously, Brites et al. (2010), through a sensory analysis on *broa*, carried out by a trained panel using open-pollinated maize populations, identified a preference, due to texture, taste, and aroma, for maize bread produced using open-pollinated populations, as opposed to maize bread produced using commercial hybrid maize varieties.

As main conclusions, the work developed under this Ph.D. opened ways in the field of participatory maize breeding in Portugal, improved the knowledge on the quality characterization of traditional maize landraces, postulating future paths for breeding these materials, and increased the basic and applied knowledge on the genetic control of quality-related traits in maize.

As final remark, and noteworthy to mention is the fact that the work developed under this Ph.D. thesis, where knowledge of the agronomic performance, quality characterization, and identification of putative genomic regions joint to develop future markers to assist selection of quality traits, is also in line with the "Goal 2: End hunger, achieve food security and improved nutrition and promote sustainable agriculture", of the 17 Sustainable Development Goals of the United Nations' 2030 Agenda for Sustainable Development (United Nations, UN, 2015).

References

Brites, C., Trigo, M. J., Santos, C., Collar, C., & Rosell, C. M. (2010). Maize-based gluten-free bread: influence of processing parameters on sensory and instrumental quality. *Food and Bioprocess Technology*, 3(5), 707-715.

Buckner, B., San Miguel, P., Janick-Buckner, D., & Bennetzen, J. L. (1996). The y1 gene of maize codes for phytoene synthase. *Genetics*, 143(1), 479-488.

Carbas, B., Vaz Patto, M.C., Bronze, M.R., Bento da Silva, A., Trigo, M.J., & Brites, C. (2016). Maize flour parameters that are related to the consumer perceived quality of 'broa' specialty bread. *Food Science and Technology*, 36, 259–267.

Chandler, K., Lipka, A. E., Owens, B. F., Li, H., Buckler, E. S., Rocheford, T., & Gore, M. A. (2013). Genetic analysis of visually scored orange kernel color in maize. *Crop Science*, 53(1), 189-200.

Chen, D., Li, Y., Fang, T., Shi, X., & Chen, X. (2016). Specific roles of tocopherols and tocotrienols in seed longevity and germination tolerance to abiotic stress in transgenic rice. *Plant Science*, 244, 31-39.

Diepenbrock, C. H., Kandianis, C. B., Lipka, A. E., Magallanes-Lundback, M., Vaillancourt, B., Góngora-Castillo, E., et al. (2017). Novel Loci Underlie Natural Variation in Vitamin E Levels in Maize Grain. *The Plant Cell* Online, tpc-00475.

Fu, Z., Chai, Y., Zhou, Y., Yang, X., Warburton, M. L., Xu, S., et al. (2013). Natural variation in the sequence of PSY1 and frequency of favorable polymorphisms among tropical and temperate maize germplasm. *Theoretical and Applied Genetics*, 126(4), 923-935.

Gao, M., Wanat, J., Stinard, P. S., James, M. G., & Myers, A. M. (1998). Characterization of dull1, a maize gene coding for a novel starch synthase. *The Plant Cell*, 10(3), 399-412.

Gwirtz, J. A., & Garcia-Casal, M. N. (2014). Processing maize flour and corn meal food products. *Annals of the New York Academy of Sciences*, *1312*(1), 66-75.

Jiang, L., Wang, W., Lian, T., & Zhang, C. (2017). Manipulation of Metabolic Pathways to Develop Vitamin-Enriched Crops for Human Health. *Frontiers in Plant Science*, 8, 937.

Jittham, O., Fu, X., Xu, J., Chander, S., Li, J., & Yang, X. (2017). Genetic dissection of carotenoids in maize kernels using high-density single nucleotide polymorphism markers in a recombinant inbred line population. *The Crop Journal*, 5(1), 63-72.

Hefferon, K. L. (2015). Nutritionally enhanced food crops; progress and perspectives. *International Journal of Molecular Sciences*, 16(2), 3895-3914.

Hellin, J., Bellon, M. R., & Hearne, S. J. (2014). Maize landraces and adaptation to climate change in Mexico. *Journal of Crop Improvement*, 28(4), 484-501.

Hoban, S., Arntzen, J. A., Bruford, M. W., Godoy, J. A., Rus Hoelzel, A., Segelbacher, G., et al. (2014). Comparative evaluation of potential indicators and temporal sampling protocols for monitoring genetic erosion. *Evolutionary Applications*, 7(9), 984-998.

Ketthaisong, D., Suriharn, B., Tangwongchai, R., & Lertrat, K. (2014). Combining ability analysis in complete diallel cross of waxy corn (Zea mays var. ceratina) for starch pasting viscosity characteristics. *Scientia Horticulturae*, 175, 229-235.

Klensporf, D., & Jelén, H. H. (2005). Analysis of volatile aldehydes in oat flakes by SPME-GC/MS. *Polish Journal of Food and Nutrition Sciences*, 14, 389–395.

Lipka, A. E., Gore, M. A., Magallanes-Lundback, M., Mesberg, A., Lin, H., Tiede, T., et al. (2013). Genome-wide association study and pathway-level analysis of tocochromanol levels in maize grain. *G3: Genes, Genomes, Genetics*, 3(8), 1287-1299.

Mendes-Moreira, P. M. R., Pego, S. E., Vaz Patto, C., & Hallauer, A. R. (2008). Comparison of selection methods on 'Pigarro', a Portuguese improved maize population with fasciation expression. *Euphytica*, 163, 481–499.

Mendes-Moreira, P. M. R., Vaz Patto, M. C., Mota, M., Mendes-Moreira, J., Santos, J. P. N., Santos, J. P. P., et al. (2009). 'Fandango': Long term adaptation of exotic germplasm to a Portuguese on-farm-conservation and breeding project. *Maydica*, 54, 269–285.

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., et al. (2016). The ensembl variant effect predictor. *Genome Biology*, 17(1), 122.

Munck, L. (2009). Breeding for Grain Quality Traits. In *Cereals* (pp. 333-366). Springer US.

Nuss, E. T., & Tanumihardjo, S. A. (2010). Maize: a paramount staple crop in the context of global nutrition. *Comprehensive Reviews in Food Science and Food Safety*, 9(4), 417-436.

Ranum, P., Peña-Rosas, J. P., & Garcia-Casal, M. N. (2014). Global maize production, utilization, and consumption. *Annals of the New York Academy of Sciences*, 1312(1), 105-112.

Revilla, P., de Galarreta, J. I. R., Malvar, R. A., Landa, A., & Ordás, A. (2015). Breeding maize for traditional and organic agriculture. *Euphytica*, 205(1), 219-230.

Romay, M. C., Millard, M. J., Glaubitz, J. C., Peiffer, J. A., Swarts, K. L., Casstevens, T. M., et al. (2013). Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biology*, 14(6), R55.

Sattler, S.E., Gilliland, L.U., Magallanes-Lundback, M., Pollard, M., & DellaPenna, D. (2004). Vitamin E is essential for seed longevity and for preventing lipid peroxidation during germination. *The Plant Cell* 16(6), 1419-1432.

Shutu, X., Dalong, Z., Ye, C., Yi, Z., Shah, T., Ali, F., et al. (2012). Dissecting tocopherols content in maize (Zea mays L.), using two segregating populations and high-density single nucleotide polymorphism markers. *BMC Plant Biology*, 12(1), 201.

Smith, M. E., Castillo, F. G., & Gómez, F. (2001). Participatory plant breeding with maize in Mexico and Honduras. *Euphytica*, 122, 551–565.

United Nations (UN) (2015). Transforming Our World: the 2030 Agenda for Sustainable Development. Resolution adopted by the General Assembly on 25 September 2015, A/RES/70/1. UN General Assembly: New York.

Vaz Patto, M.C., Alves, M.L., N.F., Almeida, C.S., Mendes, P., & Satovic, Z. (2009). Is the bread making technological ability of Portuguese traditional maize landraces associated with their genetic diversity? *Maydica*, 54(2/3), 297-311.

Vaz Patto, M.C., Mendes-Moreira, P.M., Alves, M.L., Mecha, E., Brites, C., Bronze, R., et al. (2013). Participatory plant quality breeding: An ancient art revisited by knowledge sharing. The Portuguese experience. In *Plant breeding from laboratories to fields,* S. B. Andersen (Ed.), (pp. 255–288). Rijeka, Croatia: InTech.

Vaz Patto, M.C., Moreira, P.M., Carvalho, V., & Pego, S. (2007). Collecting maize (*Zea mays* L. convar. *mays*) with potential technological ability for bread making in Portugal. *Genetic Resources and Crop Evolution*, 54(7), 1555-1563.

Wallace, J. G., Larsson, S. J., & Buckler, E. S. (2014). Entering the second century of maize quantitative genetics. *Heredity*, *112*(1), 30-38.

Wang, D., Wang, Y., Long, W., Niu, M., Zhao, Z., Teng, X., et al. (2017). SGD1, a key enzyme in tocopherol biosynthesis, is essential for plant development and cold tolerance in rice*. Plant Science*, 260, 90-100.

Wen, W., Brotman, Y., Willmitzer, L., Yan, J., & Fernie, A. R. (2016). Broadening our portfolio in the genetic improvement of maize chemical composition. *Trends in Genetics*, 32(8), 459-469.

Wheeler, T., & Von Braun, J. (2013). Climate change impacts on global food security. *Science*, 341(6145), 508-513.

Wilson, L. M., Whitt, S. R., Ibáñez, A. M., Rocheford, T. R., Goodman, M. M., & Buckler, E. S. (2004). Dissection of maize kernel composition and starch production by candidate gene association. *The Plant Cell*, 16(10), 2719-2733.

Wong, J. C., Lambert, R. J., Wurtzel, E. T., & Rocheford, T. R. (2004). QTL and candidate genes phytoene synthase and ζ-carotene desaturase associated with the accumulation of carotenoids in maize. *Theoretical and Applied Genetics*, 108(2), 349-359.

Wu, J. D., Jiang, C. P., Zhu, H. S., Jiang, H. Y., Cheng, B. J., & Zhu, S. W. (2015). Cloning and functional analysis of the promoter of a maize starch synthase III gene (ZmDULL1). *Genetics and Molecular Research*, 14, 5468-5479.

Yang, X., Yan, J., Shah, T., Warburton, M. L., Li, Q., Li, L., et al. (2010). Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. *Theoretical and Applied Genetics*, 121(3), 417-431.

Funding acknowledgement

This work was supported by Fundação para a Ciência e Tecnologia, with a Ph.D. grant (Ref. SFRH/BD/77723/2011) awarded to Mara Lisa Alves, and through the research project under the grant agreement no. PTDC/AGR-ALI/099285/2008, and through the research unit GREEN-IT (UID/Multi/04551/2013).

This work was also supported by the European Union's 7th Framework Programme and the Horizon 2020 Framework Programme under the grants agreement no. 245058 and no. 633571.