UNIVERSITÀ DEGLI STUDI DI MILANO

Dottorato di Ricerca in
Biologia Vegetale e Produttività
della Pianta Coltivata
XXVIII Ciclo

Prediction of DH lines testcross value for yield and grain moisture in maize: efficacy of QTL mapping and genome-wide approaches in multi-parental elite populations

Relatore: Dott. Salvatore Roberto PILU
Correlatori: Dott. Raffaele CAPITANIO
Dott. Nicolas RANC

Coordinatore: Prof. Pier Attilio BIANCO

Dottorando: Giovanni DELLA PORTA
Matricola: R10054

Anno Accademico 2014 - 2015

Contents

General Introduction

The global human population is expected to grow from a current 6 billion people to 9 billion by the year 2050 (United Nations Population Division, 2000). Presently, more than 10% of the world's population is undernourished. While global production of cereals (the most important food crops) has increased greatly since the 1960s, per-capita production has declined unsteadily since 1984 (Food and Agriculture Organization, 2002). Lastly, limited land availability and ongoing degradation limit the area of land available for agricultural. From 1961 to 2001, global cereal production more than doubled, from 900 million Mg to more than 2 billions (Food and Agriculture Organization, 2002). The vast majority of this growth was a result of yield (production per unit area) growth, and yield growth is the most realistic option for increasing production in the future (Gregory and Ingram, 2000; Food and Agriculture Organization, 2002) in order to be able to produce enough food for the global population.

Past yield increases have been achieved through genetic improvement in rice and wheat varieties and maize hybrids, and the modification of agricultural practices, such as the use of high levels of fertilizer, the use of pesticides, and irrigation (Borlaug, 1983; Feyerherm et al. 1988; Tollenaar, 1989; Duvick and Cassman, 1999; Khush, 1999; Reynolds et al. 1999). Future yield growth are expected to come from similar sources (Hoisington et al. 1999; Khush, 1999; Rajaram, 1999; Reynolds et al. 1999; Serageldin, 1999; Borlaug, 2000).

On a global scale, none of the three most important cereals showed a significant trend of slowing yield growth, rather, all in the last fifty years showed substantial growth at an average rate of 62, 55, and 43 kg ha^{-1} yr^{-1} for maize, rice, and wheat, respectively (Fig. 1).

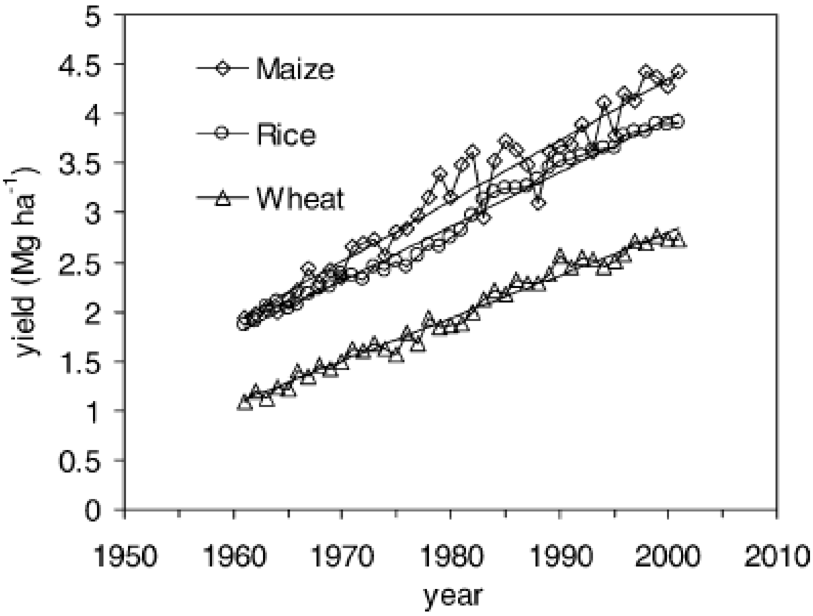


Figure 1: Global maize, wheat and rice yields from 1960 to 2001 (Hafner, 2003)

Currently maize represents the single most important cereal crop by total production; its worldwide production exceeded 875 million metric tons in 2012 (FAOSTAT 2012). The increasing importance of that crop comes also from its multiple potential uses. In many developing countries, particularly in Eastern and Southern Africa and parts of Latin America, maize is a staple food and represents the most important carbohydrate source in the human diet. In industrialized nations, on the other hand, it is primarily used as feed for livestock (in either grain or whole-plant silage form) or as industrial raw material. In recent years maize also gained importance as energy crop, particularly in the USA where it is mostly employed for the production of fuel ethanol from grain and in Europe where whole-plant biomass is used for biogas production.

Several studies in maize were conducted in order to dissect the origin of the yield gain that happened in the last 70-80 years. Of particular relevance the studies and review done by Duvick in particular for the United States maize production (Duvick, 2005); these studies give useful data and information that are valid for all the countries were that yield increase happened after hybrid introduction. Duvick tested the most important hybrids belonging to different decades in the same environments and applying different agronomic managements: the results obtained showed that the genetic improvements have been responsible for about 50-60% of the on-farm gains, and changes in cultural practices are responsible for the remainder 40-50%. These estimations have to take into account that, because breeding and management interact with each other, neither factor could have raised yields without concurrent and complementary changes in the other. Numerous estimates of genetic yield gain of maize hybrids have shown, without exception, that genetic yield gains during the past70 years have been positive and linear.

Maize grain yields in the U.S., the most important maize producer worldwide, started to rise in the late 1930s, concurrent with introduction of hybrids and improved cultural methods. On-farm yield gains averaged 115 kg ha^{-1} yr^{-1} during the years 1934-2004 (Fig.2).

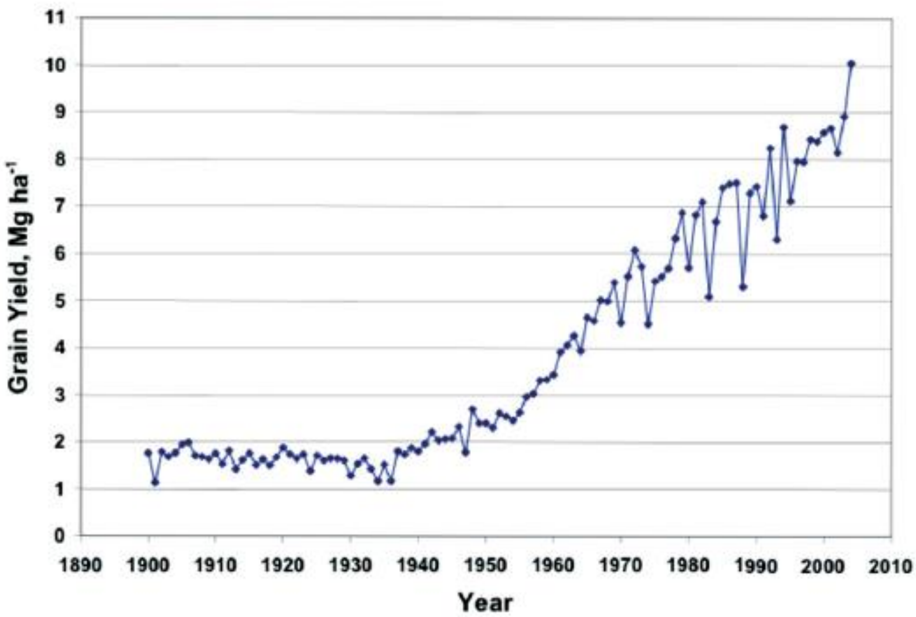


Figure 2: Average U.S. maize yields, 1940-2004 (Duvick, 2005)

One of the consequences of the hybrid introduction in order to benefit from the phenomenon of heterosis, was the development of heterotic pattern or heterotic groups: these groups were created by breeders as means of maximizing the amount of hybrid vigor and ultimately grain yield in a more predictable manner (Tracy, 2006); they were established empirically through testing and choice of lines to be recombined. Lee (1995) defines a Heterotic groups as a “collection of Germplasm which, when crossed to

Germplasm external to its group, tends to exhibit a higher degree of heterosis than when crossed to a member of its own group". Modern hybrids are the result of crossing an inbred line from a given heterotic pattern with an inbred line from a different heterotic pattern. Classification of these patterns is generally based upon several criteria such as pedigree, molecular-based associations and performance in hybrid combination (Smith, 1990).

For the Germplasm in use in North America and Europe for the FA0 400-700 maturities, typically two heterotic patterns are described (Fig.3) and used (Stiff Stalk Synthetic, SSS, and Not Stiff Stalk Synthetic, NSS) and the hybrids are made crossing two inbreds belonging to these two groups. The pedigree origin of the "Stiff Stalk" heterotic pattern traces back to a 16-line synthetic breeding population developed at Iowa State University in the 1930's by Sprague; as a group the 16 parental lines were 75% Reid Yellow Dent (Troyer, 2000). The Not Stiff Stalk heterotic pool is characterized by having a different origin from Stiff Stalk one and could be divided in three sub-groups: "Iodent", that traces back to two OPVs canned Iodent and Minnesota 13, "Lancaster", that origin from Lancaster OPV and "miscellaneous" where are included commercial hybrid derived and "maiz amargo" derived germplasms.

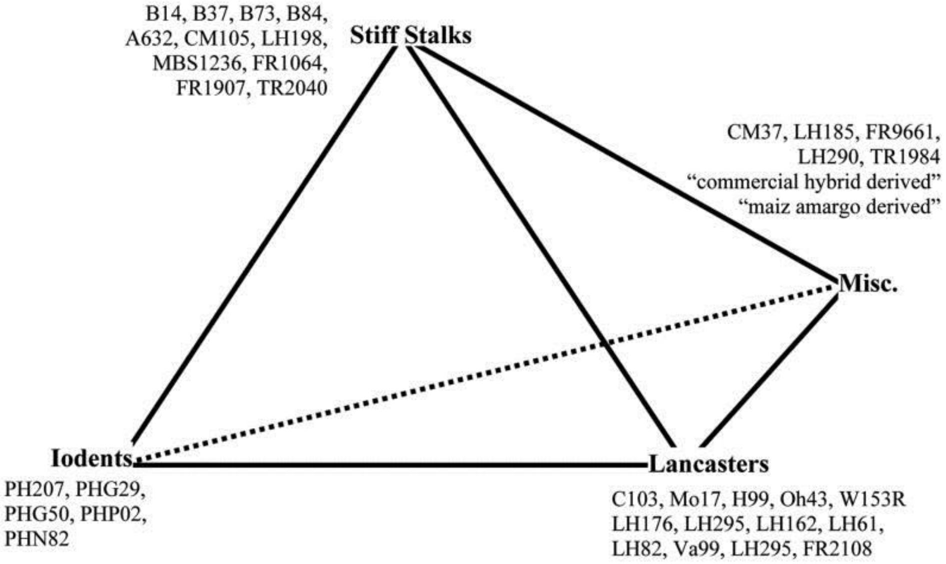


Figure 3: Northern corn belt heterotic patterns (Lee and Tollenar, 2007)

Hybrid maize traces its roots back to experiments on heterosis and inbreeding conducted by Shull and East made nearly 100 years ago, and methodology outlined by Shull (1908) gave rise to the modern hybrid maize industry (Crow, 1998). Because of the hybrid nature of the crop, modern temperate maize breeding has evolved into two very distinct activities: inbred line development and hybrid commercialization (Duvick and Cassman, 1999; Fig. 4). Inbred line development is the stage of maize breeding where the greatest amount of new genetic variation is present, created through recombination giving rise to novel alleles and new allelic combinations. In hybrid commercialization the genetic variation is potentially less, but represented by a far more refined germplasm pool; one that has been through extensive evaluation.

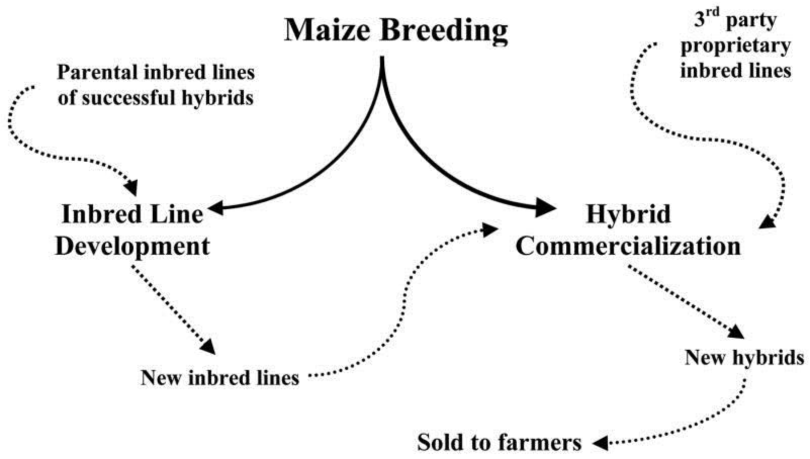


Figure 4: Overview of a modern maize breeding program (Lee and Tollenar, 2007)

The modern maize breeding program can be viewed as an open reciprocal recurrent selection program (i.e., essentially treating each heterotic pattern as a recurrently selected population) (Duvick et al. 2004). Maize breeding methodologies and philosophies have not remained the same during the hybrid era but have evolved significantly, incorporating significant scientific advances in breeding and genetic theory such as early-generation testing, rapidly adopting improvements in agronomic management practices (i.e. increased plant population densities and modern herbicide chemistries) and recognizing how best to assess genetic potential of a genotype (i.e. improvements in experimental design and data analysis). The majority of inbred development activities in temperate corn involve the use of the pedigree method of breeding (Duvick et al. 2004; Mikel and Dudley, 2006) (Fig. 5). Breeding crosses tend to be made by crossing inbred lines within a heterotic pattern and inbred lines from the other heterotic patterns are used to improve the heterotic pattern represented by the breeding cross. The typical pedigree breeding scheme generally consists of a two-parent breeding cross within a heterotic pattern. Parent selection is based on proven commercial utility of the inbred lines. An F2 population is formed from the breeding cross. Inbreeding is performed for several generations (e.g., S2) using ear-to-row with each family tracing back to different F2 plants. During the inbreeding process, genotypes with obvious defects are eliminated. Early generation testing occurs around the S2 generation, which involves forming topcross hybrids between the S2 lines and an inbred line from each of the main heterotic patterns. The resulting hybrids are evaluated in a limited number of environments and selections based on agronomic performance are made. Only S2 lines that correspond to the selected topcross hybrids will be retained in the breeding program. The selected S2 families are further inbred to the S5 generation where a second round of topcross hybrid evaluation is performed. Again an inbred line from each of the main heterotic patterns is used to form the topcross hybrids. The hybrids are evaluated in a limited number of environments and selections are based on agronomic performance compared to commercial hybrids. In general, all testing during inbred line development is done in hybrid combinations, involving relatively limited number of hybrid combinations, in relatively limited number of environments, and focused primarily on grain yield. At this stage any superior inbred lines are then considered for release to the hybrid commercialization side of the breeding activities.

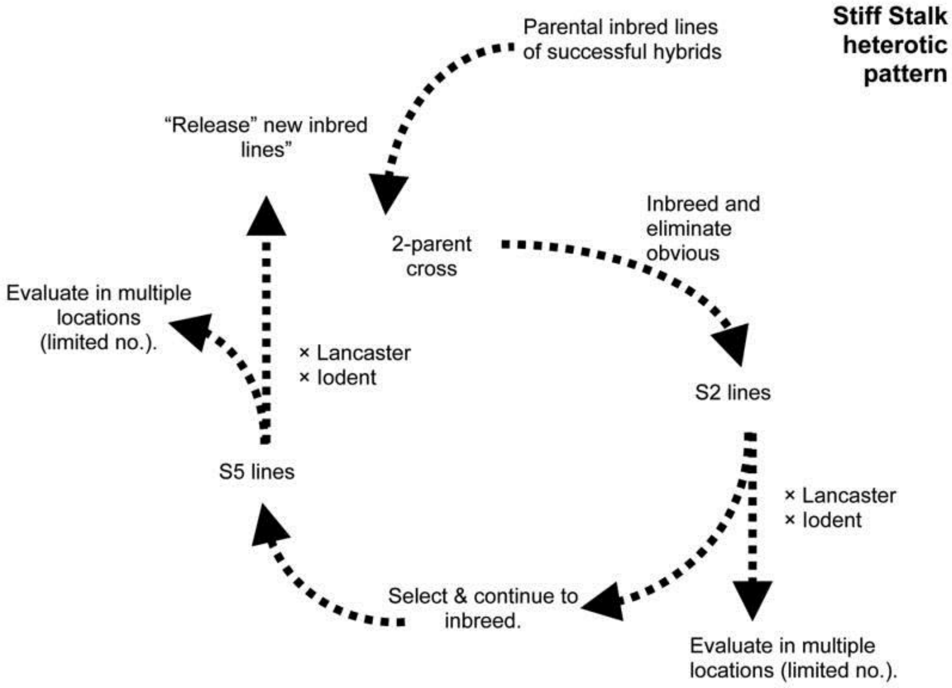


Figure 5: Typical inbred line development scheme depicting a two-parent breeding cross involving two inbred lines from the Stiff Stalk heterotic pattern (Lee and Tollenar, 2007)

Seitz identified as most important steps for modern maize breeding hybrid technology, the off-season nursery and the doubled haploid (DH) technology (Seitz 2005); in addition to that, starting from the mid-1980s, the development of abundant molecular markers, appropriate statistical procedures and user-friendly softwares, marker assisted selection became a tool available for the maize breeders (Bernardo, 2008).

The DH technology, which reduces dramatically the time necessary to obtain fully homozygous inbred lines (Prigge and Melchinger 2012), and which enables the generation of a huge number of inbred lines every year, is meanwhile worldwide routinely applied in maize breeding programs. Typically, DH lines originate from distinct crosses between related or unrelated parents. In practical breeding programs, a parent is often crossed with several other parents in a connected design, which enables the evaluation of the influence of one parent in combination with several others, related or unrelated parents. On a chromosomal level, connected designs enable the evaluation of the contribution of similar or different linkage phases on chromosomal regions, which are contributed by the parents involved.

Hybrid maize breeding involves (i) production of new candidates within each heterotic pool, (ii) evaluation of their line per se performance especially for characters related to hybrid seed production and (iii) evaluation of their testcross performance in combination with genotypes from the opposite heterotic pool (Hallauer 1990).

The implementation of molecular markers in the maize breeding (Marker-assisted selection - MAS) has been implemented widely in breeding for mono- or oligogenic resistance traits and has the potential to play an even more important role in the future. However, to date, the genetic improvement of many important polygenic resistance traits through MAS has posed significant challenges (Miedaner and Korzun 2012). Despite a large number of published quantitative trait loci (QTL) mapping studies focusing on quantitative resistance traits, very few reports demonstrate the successful application of QTL-based MAS in a practical breeding program

(St. Clair 2010). In fact, the long breeding history in cereals suggests that if any major QTL for grain yield were present to begin with, then the favorable alleles at these major QTLs would have been fixed during the domestication process or during previous selections, so currently much of the variation is controlled by many QTL with small effect (Bernardo, 2008). Typical populations used for QTL mapping include F2, backcross (BC) or recombinant inbred (RI) populations derived from only two parents (also referred as bi-parental population). The limitations of using such bi-parental populations are that only two alleles are analyzed and that genetic recombination in these populations is limited which limits the resolution for QTL detection. In recent year, to encompass this limitation, multi-parent cross designs are suggested in order to significantly increase mapping resolution and power. This approach is a bridge between bi-parental population and Association mapping on inbred panels approach. Multi-parental populations are produced by crossing more than two inbred lines applying different schemes, as for example Multi-parent Advanced Generation Inter-Crosses (MAGIC) populations, Nested Association Mapping (NAM) panel, multi-parental populations (Dell'Acqua, 2015).

Taking into account the current widespread adoption of the double haploid (DH) technology in major crops such as maize - that led to a tremendous increase in the number of new lines produced every year in the public and private sector during the past decade – and the development of very dense molecular marker coverage of the genome and more efficient software for marker trait association software, highly effective selection schemes involving the use molecular markers should be developed to the most promising lines and hybrid combinations as early and as efficiently as possible (Technow et al. 2014). Taking into account the almost stable cost of phenotypic yield testing and the more and more cheaper cost of molecular markers analysis, a molecular marker based prediction of untested new DH lines is one of the most promising approach in that direction, if adequate levels of prediction accuracy are reached.

Among the major trait of interest for the maize breeders, the yield and the harvest moisture represent two key performance traits and are well known as quantitative traits controlled by a very large number of genes, it looks interesting and not already explored in the literature a comparison of the main molecular marker based prediction methods: QTL mapping and Genome-Wide prediction.

Grain yield is the primary trait of interest for the maize breeders; testing for grain yield is always done in hybrid combinations and testing is done using the most common current agronomic practices. The specific trait that is assessed is typically the machine harvestable grain yield adjusted to 15.5% grain moisture; this is because that moisture level is considered the water content is compatible with good seed preservation during storage (Capelle et al. 2010).

Grain moisture at harvest (%) is used by many breeders as the best assessment of the “maturity” of a given hybrid, that refers to whether a hybrid is adapted to a particular environment. The “ideal” hybrid is one that maximizes the full growing season available reaching the physiological maturity (i.e. black layer formation) before the first killing frost and facilitate moisture loss from the kernel (i.e. dry down).

If grain moisture increases, shelling efficiency and grain quality will be reduced and drying costs and shrink age penalties increase. Grain moisture content is also an important factor which impacts the fungal development of Ear rot species (Xiang, 2012). Ear rot is one of the most prevalent ear diseases of maize occurring worldwide and is mainly caused by Fusarium species and results in reduced grain yield but the main loss from Ear rot is due to the contamination of the grain with mycotoxins which are a threat to the safety of both humans and livestock (Pestka 2007; Voss et al. 2007).

There is a strong positive linear relationship between grain moisture and grain yield: longer season hybrids are higher yielding and have higher grain moisture; farmers look to the best equilibrium between the two traits for their specific environmental and agronomic conditions.

These two traits showed different levels of heritability, with grain yield among the lowest of all traits of interest for maize (<30%) and grain moisture with medium-high heritability level (between 50 and 70%) (Hallauer and Miranda, 1988).

Both traits are well known as quantitative traits, with a very large number of genes determining it. In particular yield encompass all the genes that determine the fitness of the organism, because depends on all the alleles that positively affect all the pathways and functions affecting seed production and also to all those that influence to resistance to the biotic and abiotic stresses that can result in stalk or root lodging or dropped ears.

Recent advances in marker genotyping technologies, coupled with new and powerful statistical methods, allowed the development of MAS towards genome-wide selection (Meuwissen et al. 2001). This approach differs from traditional QTL-based MAS in its ability to exploit information provided by dense genome-wide single nucleotide polymorphism (SNP) markers, which are used to predict the total genetic value of genotypes (genome-wide prediction, GP). Statistical methods making use of information from all available SNP markers are able to cover a large number of small genetic effects and should be suitable for highly polygenic traits (de los Campos et al. 2013). In addition, GP has been shown to capture adequately large effect QTL and additionally cover the remaining genomewide effects in a single statistical model (Wimmer et al. 2013).

References

Bernardo, R. (2008) Molecular markers and selection for complex traits: Learning from the last 20 years. Crop Sci. 48: 1649-1664

Borlaug, N.E. (1983) Contributions of conventional plant breeding to food production. Science 219, 689–693

Borlaug, N.E. (2000) Ending world hunger. The promise of biotechnology and the threat of antiscience zealotry. Plant Physiol. 124, 487–490

Capelle V, Remoué C, Moreau L, Reyss A, Mahé A, Massonneau A, Falque M, Charcosset A, Thévenot C, Rogowsky P, Coursol S, Prioul JL (2010) QTLs and candidate genes for desiccation and abscisic acid content in maize kernels. BMC Plant Biol. 2010 Jan 4;10

Crow, J.F. (1998) 90 years ago: The beginning of hybrid maize. Genetics 148:923–928

Dell'Acqua M, Gatti DM, Pea G, Cattonaro F, Coppens F, Magris G, Hlaing AL, Aung HH, Nelissen H, Baute J, Frascaroli E, Churchill GA, Inzé D, Morgante M, Pè ME (2015) Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in Zea mays. Genome Biol. 2015 Sep 11;16:167

Duvick, D.N., and K.G. Cassman. (1999) Post–Green Revolution trends in yield potential of temperate maize in the north-central United States. Crop Sci. 39:1622–1630

Duvick, D.N., J.C.S. Smith, and M. Cooper. (2004) Long-term selection in a commercial hybrid maize breeding program. Plant Breed. Rev. 24:109–151

Duvick, D.N. (2005) Genetic progress in yield of United States Maize (Zea mays L.). Maydica 50: 193-202

Feyerherm, A.M., Kemp, K.E., Paulsen, G.M. (1988) Wheat yield analysis in relation to advancing technology in the Midwest United States. Agron. J. 80 (6), 998–1001.

Food and Agriculture Organization (2012) Internet database: http://www.fao.org

Gregory, P.J., Ingram, J.S.I. (2000) Global change and food and forest production: future scientific challenges. Agric. Ecosyst. Environ. 82, 3–14

Hafner S (2003) Trends in maize, rice, and wheat yields for 188 nations over the past 40 years: a prevalence of linear growth. Agriculture, Ecosystems and Environment 97: 275–283

Hallauer, A.R. and Miranda J.B. (1988) Quantitative Genetics in Maize Breeding, 2nd ed. Iowa State University Press, Ames, IA.

Hoisington, D., Khairallah, M., Reeves, T., Ribaut, J., Skovmand, B., Taba, S., Warburton, M. (1999) Plant genetic responses: what can they contribute toward increased crop productivity? Proc. Natl. Acad. Sci. USA 96, 5937–5943

Khush, G.S. (1999) Green revolution: preparing for the 21st century. Genome 42, 646–655

Lee, M. (1995) DNA markers and plant breeding programs. Adv. Agron. 55: 265-344

Lee, E.A., and Tollennar, M. (2007) Physiological basis of successful breeding strategies for maize grain yield. Crop Sci. (S3):S202-S215

Meuwissen, T. H. E., Hayes B. J., and Goddard M. E. (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819–1829

Miedaner, T., Korzun, V. (2012) Marker-assisted selection for disease resistance in wheat and barley breeding. Phytopathology;102(6):560-6 Review.

Mikel, M.A., and J.W. Dudley (2006) Evolution of North American dent corn from public to proprietary germplasm. Crop Sci. 46:1193–1205

Pestka, J.J. (2007) Deoxynivalenol: toxicity, mechanisms and animal health risk. Anim Feed Sci Tech 137:283–298

Prigge, V., Melchinger A. E. (2012) Production of haploids and doubled haploids in maize, Plant Cell Culture Protocols, Ed. 3, edited by Loyola-Vargas V. M., Ochoa-Alejo N., editors. Humana Press–Springer Verlag, Totowa, NJ

Rajaram, S. (1999) Approaches for breaching yield stagnation in wheat. Genome 42, 629–634

Reynolds, M.P., Rajaram, S., Sayre, K.D. (1999) Physiological and genetic changes of irrigated wheat in the post-green revolution period and approaches for meeting projected global demand. Crop Sci. 39, 1611–1621

St Clair DA. (2010) Quantitative disease resistance and quantitative resistance Loci in breeding. Annu Rev Phytopathol. 48:247-68. Review

Seitz, G. (2005) The use of double haploids in corn breeding. In: Proc. of the 41[th] Annual Illinois Corn Breeder's school 2005. Urbana-Champaign, Il, USA, p. 1-7

Serageldin, I. (1999) Biotechnology and food security in the 21[st] century. Science 285, 387–389

Shull, G. H. (1908) The composition of a field of maize. J. Hered. os-4: 296–301

Smith, O.S., Smith, J.S.C., Bowen, S.L., Tenborg, R.A., Wall, S.J., (1990) Similarities among a group of elite maize inbreds as measured by pedigree, F1 grain yield, grain yield, heterosis and RFLPs. Theor. Appl. Genet. 80: 833-840

Technow, F., Schrag, T.A., Schipprack, W., Bauer, E., Simianer, H., Melchinger, A.E. (2014) Genome properties and prospects of genomic prediction of hybrid performance in a breeding program of maize. Genetics197(4):1343-55

Tollenaar, M. (1989) Genetic improvement in grain yield of commercial maize hybrids grown in Ontario from 1959 to 1988. Crop Sci. 29, 1365–1371

Tracy, W.F., Chandler, M.A. (2006) The historical and biological basis of the concept of heterotic patterns in corn belt dent maize. In: Lamkey KR, Lee M (eds) Plant breeding: The Arnel R Hallauer international symposium. Blackwell Puslishing, Ames, IA, pp 219-233

Troyer, A.F. (2000) Origins of modern corn hybrids. In: Wilkinson D (ed), Proceedings of the 55th annual corn and sorghum research conference. Am Seed Trade Assn, Washington DC, 27-42

United Nations Population Division (2000) World population prospects the 2000 revision highlights. Population Division, Department of Economic and Social Affairs, United Nations, New York, NY

Voss, K.A., Smith, G.W., Haschek, W.M. (2007) Fumonisins: toxicikinetics, mechanism of action and toxicity. Anim Feed Sci Tech 137:299–325

Wimmer, V., C. Lehermeier, T. Albrecht, H.-J. Auinger, Y. Wang et al. (2013) Genome-wide prediction of traits with different genetic architecture through efficient variable selection. Genetics 195: 573–587

Xiang, K., Reid, L.M., Zhang, Z.M., Zhu, X.Y., Pan G.T. (2012) Characterization of correlation between grain moisture and ear rot resistance in maize by QTL meta-analysis. Euphytica 183:185–195

Article Type: Original Article in preparation

Prediction of DH lines testcross value for yield and grain moisture in maize: efficacy of QTL mapping and genome-wide approaches in multi-parental elite populations

Giovanni Della Porta[1], Raffaele Capitanio[1], Filippo Geuna[3], Frederic Cossic[2], Javier Betran[2], Nicolas Ranc[2], Roberto Pilu[3],.

[1] Syngenta Italia S.p.A., Via Per Soresina, 26020 Casalmorano (CR), Italy

[2] Syngenta France S.A.S., 12 chemin de l'Hobit 31790 Saint Sauveur, France

[3] Dipartimento di Scienze Agrarie e Ambientali - Produzione, Territorio, Agroenergia, Università degli Studi di Milano, Via Celoria 2, 20133 Milano, Italy

Corresponding author:

R. Pilu
Dipartimento di Scienze Agrarie e Ambientali - Produzione, Territorio, Agroenergia, Università degli Studi di Milano
Via Celoria 2, 20133 Milano, Italy
e-mail : salvatore.pilu@unimi.it
Fax number: + 39-2-50316521
Phone number: + 39-2-50316549

Running Title: QTL mapping and genome-wide approaches using multi-parental elite DH populations

Abstract

Grain Yield (GY) and Grain Moisture content at harvest (GM) are complex quantitative traits and by far the two performance traits of major interest for maize breeder. Prediction accuracy of the testcross performance of untested Double Haploid (DH) maize lines for these two traits is of tremendous importance in order to increase genetic gain. We analysed genomic and phenotypic data of testcross progenies of 1066 DH lines genotyped with 3072 SNP markers and derived from three large half-sib populations phenotyped for GY and GM in eight locations and applied cross-validation method to compare the accuracy of whole genome prediction (GP) and QTL based prediction approaches.

GP showed higher accuracy for both traits with mean predictive ability of 0.58 and 0.73 for GY and GM respectively in comparison of 0.15 and 0.30 as mean predictive ability obtained using the QTL-based approach. Both methods showed higher accuracy in predicting GM in comparison to GY. The lack of accuracy for QTL based prediction confirmed the major issues traditionally faced in playing with QTL for polygenic and complex traits. For GP the simultaneous use of the three half-sib population did not increase the accuracy prediction obtained working within the same parental population while the predictive ability dropped when predicting a population with a training set formed by other population (0.39 and 0.58 versus 0.53 and 0.73 for GY and GM respectively) confirming the strong influence of genetic relationship between estimation set and test set on the predictive ability. Our experimental results confirm that Whole Genome prediction accuracy is surpassing QTL prediction accuracy for the two maize-breeding target-traits and raised the concern on the way grain yield and grain moisture QTL had been implemented in breeding program so far. One could be interested in exploring possibility to combine relevant QTL and whole genome prediction together to advance towards performance prediction accuracy increase.

Key words: *Zea mays*, grain yield, grain moisture, DH lines, QTL, genomic selection, multi-parental populations,.

Introduction

Maize breeders work to deliver higher yielding germplasm; in addition to Grain Yield (GY) they typically use Grain Moisture content at harvest (GM) as key selection trait: the combination of the characteristics for these two traits is used as key parameter for making selections among the different genotypes in testing (Hallauer and Miranda 1988; Meuwissen et al. 2001; Bekavac et al. 2008; Sala et al. 2006; Sala et al. 2012). The importance of the GM trait is due to the fact that maize is intolerant to cold stress being a native of subtropical areas and for this reason is mainly cultivated in the temperate zones (from mid to short-season areas) where it is planted in the spring. Its capacity to reach the physiological maturity and low relative humidity at harvest before the bad weather of Autumn is very important, because it has an impact on GY, grain quality, in particular because the correlation between GM and ear rot diseases (Xiang et al. 2012) and on the profitability of the crop because of additional drying of grain prior utilization by using fossil fuels.

Maize is currently widely diffused around the world and it is considered as a tradable commodity used for several purposes such as to feed animal, for the agroindustry, to produce energy and in direct human consumption. This success was due by hybrid breeding pioneered in maize by Shull in 1908 and then used on several vegetable crops although genetic mechanisms involved in the hybrid vigour are still unknown (Duvick 1999; Silva Dias

2010; Lippman and Zamir 2006). This technology is based on the development of parental inbred lines that, when crossed among them, lead to hybrid performance; selection of the best combinations is the basis to produce commercial hybrids. A weakness of this approach is that the yield performance and selection of new inbred lines are poorly indicators of the performance of derived hybrids (Bekavac et al. 2008; Hallauer and Carena 2009). Hence, in maize breeding programs the genetic value of new inbred lines is assessed by their testcross performance with testers from the opposite heterotic pool in replicated multi-environment yield trials.

With the advent of double-haploid (DH) technology in maize, fully homozygous inbreds lines can be generated rapidly, at low cost and in great numbers (Wedzony, 2009) and this leads to a vast expansions of potential hybrids that can be generated using the large number of DH lines entering each year a modern breeding program; because producing and testing all these hybrids is impossible, models of predictions of their performance is of tremendous importance (Bernardo, 1996).

Until recently, prediction of testcross value of untested lines has not played an important role in plant breeding but now the availability of large genotypic information at progressively lower cost and the development of more sophisticated analysis software could allow reaching higher level of prediction accuracies.

In addition to the classical marker-assisted selection (MAS), where only a subset of significant markers, linked to mostly large-effect QTL, is used for selection, the Genomic Prediction (GP) approach, originally developed in animal breeding, where all available markers are considered without significance test in order to capture and used to predict the genotypic performance of a given genotype, is suggested as eligible to be incorporated into plant breeding programs (Meuwissen et al, 2001; Technow, 2014, Lehermeier, 2014).

Because both GY and GM traits are well known as highly polygenic traits and controlled by several small effect genomic region, it has been validated through simulation that genomewide selection that exploits cheap and abundant molecular markers is superior to Marker Assisted Recurrent Selection (Bernardo et al. 2006). Use of molecular marker to support germplasm performance prediction also gain in popularity due to the continuously decreasing cost of genotypic data compared to phenotypic data (Jannink et al. 2010).

The accuracy of marker based prediction is most often evaluated by applying cross-validations studies, where all genotypes are randomly divided into training and validation sets. The training set is used to train the prediction model and estimate the marker effect and the accuracy of the genomic prediction model is evaluated by comparing the predicted with the observed values in the validation set (Zhao, 2013). The composition of the training set is one of the key parameters affecting the GP accuracy (Abera Desta, 2014) and generally sample size and relatedness between Training Set and Validation Set has a positive effect on the accuracy.

In our study we addressed the above-mentioned questions in a comparative study of the QTL-based prediction and Genomic Prediction approaches on a real dataset coming from breeding program.

We investigated the efficiency of QTL based and GP approaches for two traits of major interest for maize breeders (GY and GM) in an elite multi-parental DH maize population sharing one common parent, using a cross validation method. Our population represent the typical situation in a maize breeding program where one pillar inbred is crossed with several others in order to get improved and develop better new cycle inbred lines. The specific objectives were to use real data from breeding population to (i) compare the level of prediction accuracy of the QTL-based in comparison to the GP prediction, (ii) asses the GP predictive ability gain coming from combining the half-sib population together in comparison to the prediction within the same bi-parental population and (iii) asses how the GP predictive ability for a given population is affected by the presence or the absence of its progenies in the training set.

Materials and methods

Plant material

Four FAO 6-700 elite inbreds originating from Non Stiff Stalk heterotic pool were used as parents of three large DH populations following this scheme: 3 inbreds (L1, L2, L3) were crossed to the same inbred line (L4) for the development of 3 connected populations sharing L4 as common parent (half-sib populations): Pop1 (L1 x L4), Pop2 (L2 x L4), Pop3 (L3 x L4) comprising 421, 388 and 257 DHs respectively for a total population dimension of 1066 DH lines. This multiparental scenario represents the typical situation in a maize breeding program where one pillar inbred is crossed with several others in order to get improved and develop better new cycle inbred lines.

DH lines were developed using *in vivo* haploid induction technology (Röber et al. 2005). As expected the success of DH line production varied for the three populations so leading to differences in population size despite the same number of kernels per breeding starts sent to the DH process. In the 2013/2014 winter season testcrosses were produced in winter nursery by crossing each DH line with one inbred used as common tester (L5) belonging to the opposite heterotic pool (Stiff Stalk Synthetic) in order to produce hybrid seeds to be tested in field trials in the next summer season. The L5 was used as seed parent in hybrid seed production because belonging to the heterotic pool that typically has this role in commercial seed production and in order to produce higher quality and more uniform hybrid seed. Seeds from the testcrosses were used in field trials. All genetic material was proprietary and supplied by Syngenta Italia Spa, Casalmorano, Italy.

Field experiments and traits measured

Field trials were conducted in eight locations located in key South Europe corn growing areas (six in Italy-Po valley and two in Spain). In all the locations the best agronomic practices were applied as done by local famers (including full irrigation and complete control of weeds). For each population a separate experiment was created using Randomized Complete Block as experimental design, with one replication by location. Hybrids were machine planted in a 2 rows plot 6m length and machine harvested; Grain Yield, GY, (q ha^{-1} at standard 15.5% moisture) and Grain moisture, GM, (%) were collected at harvest time using an experimental combine. Current best commercial hybrids sold in that areas were included as check reference and planted in several replications in each locations to asses field spatial variation: one commercial hybrid has been repeated 312 times; other commercial hybrids have been repeated 120, 64 and 32 times across the experiments. The coefficient of variation (cv) for GY was measured for the repeated checks planted in each location.

Marker analysis and linkage maps

DH populations (N = 1066) and their parents were genotyped with 3,072 single nucleotide polymorphism (SNP) markers, distributed evenly across the genome, using two custom Illumina GoldenGate SNP arrays (Illumina Inc., San Diego California, USA).

The SNPs represented a mixture of a subsample of the Illumina MaizeSNP50 BeadChip (Ganal et al. 2011) and internal Syngenta SNP chip. Out of the 3,072 original SNP markers, a subset of high-quality SNPs, polymorphic in at least one of the populations, was selected according to the following criteria: (i) a call rate higher than 0.90, (ii) a minor allele frequency higher than 0.05 and (iii) less than 20 % missing values. After these quality checking steps, 1,164 SNPs were available across the three populations for further analysis. DH lines with more than 20 % missing data in these 1,164 SNPs were discarded, thereby leaving a total of N = 941 DH lines for further analysis, with 352, 357 and 232 DH lines for Pop-1, Pop-2 and Pop-3 respectively. For each marker and population, deviations from the expected segregation ratio were tested with a Chi-square test using the sequentially rejective Holm–Bonferroni method (Holm 1979). The marker-based genetic distance between the four parental lines was calculated using the Modified Roger distance (Rogers 1972). Linkage maps were constructed individually for each population by using a maximum likelihood mapping approach and Haldane's mapping function (Haldane 1919). The consensus map was then calculated using the 1,164 SNPs that were polymorphic in at least one of the three populations. All linkage maps were constructed with JoinMap version 4.1 software (Van Ooijen, 2006).

Statistical analyses of phenotypic data

Location coefficient of variation (cv) for grain yield has been calculated for the repeated checks planted in each location in order to evaluate the field uniformity.

Analyses of variance across environments were performed using R version 3.2.0 software (R Core Team, 2013). Fixed linear models have been used to estimate variance components and significance of estimated effects:

$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ij}$

With:

Y_{ij} the phenotypic observation genotype i on location j,

μ the overall mean,

α_i the fixed effect of genotype i,

β_j the fixed effect of location j,

$\alpha\beta_{ij}$ the fixed interaction of genotype i on location j

and ε_{ijk} the random residual error.

The genotypic-mean heritabilities have been computed from Expected Mean Square estimated from the fixed effect linear model, according to Hallauer, Carena & Miranda Filho formula (2010):

$$\hat{h}^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_e^2/_{re} + \hat{\sigma}_{ge}^2/_{e} + \hat{\sigma}_g^2}$$

With:

r the number of repetition

e the number of locations

$\hat{\sigma}_g^2$ the estimated genotypic variance

$\hat{\sigma}_{ge}^2$ the estimated genotype*location interaction variance

$\hat{\sigma}_e^2$ the estimated residual variance.

A Mixed Linear Model, using lme4 package, has been used to extract the Best Linear Unbiased Estimator of genotypes that has been later used in both QTL detection and Genomic Selection analysis.

We used the following model:

$Y_{ijk} = \mu + \alpha_i + \beta_j + b_{kj} + \alpha\beta_{ij} + \varepsilon_{ij}$

With:

Y_{ij} the phenotypic observation genotype i on location j,

μ the overall mean, α_i the fixed effect of genotype i,

β_j the effect of location j,

b_{kj} the random effect of experiment k within location j,

$\alpha\beta_{ij}$ the random interaction of genotype i on location j

and ε_{ijk} the random residual error.

QTL mapping

QTL analyses were based on the genetic consensus map and were performed combined across all three populations (hereafter referred to as joint-population QTL analyses), using the MCQTL V5.5.6 software package (Jourjon et al. 2005) to encompass the limitations in terms of QTL detection coming from using only bi-parental populations. We used forward stepwise regression along with the iQTLm method (Charcosset et al. 2001). For each trait and populations empirical LOD thresholds at the 0.05 genome-wide significance level were assessed from 1,000 permutations, according to Churchill and Doerge (1994). LOD support intervals of QTL positions were defined as the map distance in cM spanning a LOD drop of one unit on each side of the LOD peak. QTL were defined as colocalizing if their respective peaks where 5 cM apart. A connected additive QTL model was implemented and in our specific case, the connected model estimated four allelic effects at each QTL (the effects of L1, L2, L3 and L4 parental alleles) in which the effect of the common parent L4 was assumed to be the same in all three populations. The total proportion of variance explained by the model (R^2) and the proportion of variance explained by individual QTL were calculated according to Mangin et al. (2010).

Genome-wide prediction

Genomic prediction framework provided by the Synbreed R package (Wimmer et al. 2012) was used for all genome-wide analysis.

All polymorphic SNP markers meeting quality criteria (N = 1,164) were used in the genome-wide prediction of Grain yield at standard moisture (GY) and Grain Moisture at harvest (GM). Marker genotypes were coded A or B, A being alleles of L4 (the common parent between the three populations) and B the alleles of other parent of the populations. Missing marker genotypes were imputed using the “family” procedure from the Synbreed R package. A genomic best linear unbiased prediction (GBLUP) model was used to predict the genetic values of DH lines. The realized relationship matrix between the DH lines of the three populations was computed based on marker data according to the method proposed by Habier et al. (2007).

Cross-validation for QTL and genome-wide prediction models

The prediction performances of QTL-based and GP models were compared in a *joint-population* framework, taking into account all the three populations simultaneously. For both QTL and GP analyses, we performed twofold CV: the data set was split into 2 subsets, one subset comprising 50 % of the DH lines built the estimation set (ES) and was used for model training, whereas the remaining subset (50 % of the DH lines) constituted the test set (TS). This is the typical situation in maize breeding programs where a part of the available DH lines created, used as ES, are genotyped and phenotyped and the remaining ones are only genotyped and used as TS and their value for the traits of interest is predicted using available marker data. The process was replicated five times with varying allocations of DH lines to the two CV subsets. Each ES and each TS comprised DH lines from all three populations. For each of the five CV subsets, the predictive performance of the QTL-based and GP models were evaluated.

For the QTL model, predictions of DH lines in the TS were based on the sum of additive effects of all significant QTL detected in the ES, whereas GBLUP predictions were based on the effects of all polymorphic SNP markers estimated in the ES. The *cvMCQTL* R package (Wimmer, 2012) was used to make the CV process, by running a CV loop on the QTL mapping routine of MCQTL.

Predictive abilities of the different models were calculated as Pearson's correlation coefficients between predicted and observed trait values in each TS. An overall mean predictive ability with standard deviation was calculated according to Luan et al. (2009).

GP was evaluated also in additional CV scenarios: (i) prediction across biparental populations (hereafter referred to as *across-population prediction*) and (ii) prediction within biparental populations (hereafter referred to as *within-population prediction*). In *across-population prediction*, the ES comprised a merged data set of two populations, whereas the remaining population(s) represented the TS; in within-population prediction the ES comprised a data set of a single population, and the remaining part of that population represented the TS.

Results

Phenotypic analysis

Using the repeated checks data, the Coefficient of Variation for yield was calculated and it ranged from 7% (in Loc-1 and Loc-5) to 12% in Loc-7, indicating a good field uniformity conditions in all the testing locations (Supplementary 1). Because of good phenotypic data quality, no any spatial correction was applied.

Grain yield ranged from 50.8 to 192.2 q ha^{-1}. Pop-1 and Pop-3 had the highest and the lowest average with respectively 134.8 and 130.8 q ha^{-1}. For GM the data ranged from 9.6 to 32.5%. Pop-1 and Pop-2 had the highest and lowest values with respectively 18.9 and 18.1% (Table 1).

Boxplots for the two traits across locations are presented in Figure 1. For GY the variation across locations was smaller than for GM. For GY the highest yield location was Loc-2 with 145.1 q ha^{-1} on average and the lowest one was Loc-3 with 117.3 q ha^{-1}. The average grain yield in 7 out of the 8 locations was comprised between 135 and 145 q ha^{-1}. For GM trait, Loc-7 and Loc-8, that were the two locations planted in Spain, showed significantly lower harvest moisture level (12.1 and 11.3% respectively) than the remaining ones; the harvest moisture of all the locations planted in Italy (Loc-1 to Loc-6) was comprised between 16.1 and 26.7% .

For both Grain Yield and Grain Moisture, the overall ANOVA showed highly significant genotype effect (p-value < $2.2e^{-16}$), location effect (p-value < $2.2e^{-16}$) and genotype*location (p-value = $2.93e^{-10}$ and $1.94e^{-12}$, respectively for GY and GM) (Supplementary 2).

Heritabilities (h^2) were calculated from results of ANOVA table and were equal to 0.75 and 0.89 respectively for GY and GM, confirming the good quality of phenotypic data collected (Supplementary 3).

BLUEs calculated data are showed in supplementary 4.

Marker analysis and genetic maps

The overall number of polymorphic SNP markers across populations was 1164, and 586 of these SNPs were polymorphic in all three populations. The three populations showed similar number of segregating SNPs (1,146, 1,062 and 1049 respectively for Pop1, Pop2 and Pop3); this is in accordance with the similar level of genetic distance of the L1, L2 and L3 parents in comparison with L4 common parent (ranging from 0.59 to 0.61) (Supplementary 5). The consensus map across all three populations included 1164 informative SNPs and displayed a total length of 1631cM over the ten chromosomes (Figure 2).

The relationship matrix shows high level of intra-population relatedness and still some consistent inter-population genomic covariance (Figure 3). This dataset really represent breeding population where several populations shared a common elite parent with sometimes a low level of genetic variance between populations.

QTL mapping

The results of the *joint-population* QTL analysis are presented in Table 2. Nine and Fourteen QTL were identified for traits GY and GM traits respectively. R^2 values for individual QTL varied between 1.91 and 11.50 for GM and between 1.91 and 14.10 for GY, Two QTL for GY (1 on chromosome 3 and one on chromosome 5) co-localized with QTL for GM. All other QTL for GM trait did not co-localize with GY ones. Additive allelic effects contributed by parents L1, L2, L3 and L4, estimated from a connected QTL model, are indicated in the Table 2. As expected, there is no one predominant parent consistently contributing to lower or higher trait but all four contributed in both directions depending on QTL and trait.

Predictive abilities of QTL-based and genome-wide prediction models

The prediction performance of QTL-based and GBLUP models was compared using twofold CV. Table 3 presents mean predictive abilities of the two approaches from *joint-population prediction* scenarios. Mean predictive abilities of GBLUP were consistently higher than the corresponding mean predictive abilities of the QTL-based model, 0.59 versus 0.15 and 0.73 vs 0.28 respectively for GY and GM. The higher heritability of GM when compared to GY resulted in higher predictive abilities of both the GBLUP and QTL models. This can be explained by a simpler genetic determinism with less causative genetic elements explaining trait variance for GM compared to GY.

GP within and across-population prediction

For the traits GY and GM, results from GP *across-population* and *within-population prediction* scenarios are summarized in Table 4. Mean predictive abilities were very similar for both traits and ranged between 0.39 and 0.58 for the two different sampling methods. The *within-population prediction* means were in general significantly higher than these measured for *across-*

population predictions, as expected taking into account the positive correlation of genetic similarity between TS and ES and the accuracy of the prediction. Although, higher predictive abilities could be expected because of strong relationship between parents and connection of the populations with a common parent. For both sampling method, the prediction means for GY were significantly lower than those for GM, in accordance with the different heritability measured for the two traits.

Discussion

Grain Yield and Grain Moisture are the two main traits driving decision making during maize breeding hybrid promotion and advancement. Those traits are mainly assessed during harvest using experimental combines: this implies a lot of resources for field trialing and this is particularly true if high quality experimentation is targeted (number of locations, repeated experimental design, number of testers). This also implies short timeline for decision making between harvests and planning of winter nursery experiments for seed increase and testcross on different testers for next year field testing.

Marker Assisted Breeding has been surveyed for many years now to decrease timeline for material creation but also to increase precision of information that is manipulated (Moreau et al. 1998). Genetic Gain is expected to drastically increase with molecular information management because impacting both accuracy of material performance and timeline (Heffner et al. 2009). Genomic Selection (GS), now, afford huge expected impact as it was observed on dairy cattle were adoption of GS can save up to 92 % of breeding cost (Hayes et al. 2009).

Our study aimed at comparing the expected efficiency of both QTL or whole genome based prediction of unobserved genotypes for the two most important traits targeted in maize breeding program.

QTL detection of joint populations.

The QTL detection done through a joint-population method allowed detection of important QTLs for both Grain Yield and Grain Moisture. Some of those QTLs showed strong LOD values with important related effects. The joint-population algorithm implemented in MCQTL software allows estimating each parental effect for any QTL due to the connected model (Mangin et al. 2010). The parents carrying the favourable alleles change according to the QTL even if we can identify a consistent ranking of parents when summing the different QTL additive effect values. For example, L4 has the main overall additive effect for Grain Yield QTLs which is consistent with its optimal breeding value (Table 2). The number of overall QTLs for GM was greater compared to GY which is not expected looking at heritability of traits but overall R^2 fit with heritabilities values. The adjusted mean across locations used for QTL location can explain the limited number of QTLs for Grain Yield because phenotypic variance specific to single location is lost. The QTLs could be directly implemented on Marker Assisted Recurrent Selection (Charmet et al. 1999) with aim to combine alleles from different parents into a single ideotype. QTLs that explain main part of phenotypic variance could be also isolated in Near Isogenic Lines to validate their effects. Definitely, the QTL based predictive ability could be improved by integrating QTL*Environment interaction into an integrative crop model to predict adaptation of allelic combination into different scenarios (Tardieu and Tuberosa, 2010, Cooper et al. 2009),

QTL based prediction of unobserved genotypes.

The predictive ability assessed through cross validation of QTL detection on the joint populations was possible using cvMCQTL package (Wimmer 2012). The two-fold cross validation has been preferred compared to five-fold proposed by Foiada et al. (2015) to better represent the size of populations used internally for QTL detection. The different values of predictive abilities were, in average, higher for GM compared to GY (Table 3). This is expected since heritability and QTL R^2 was higher for the former trait. The values of QTL based predictive ability found in this study were lower compared to the one reported by Foiada et al. (2015) even if we manipulated traits with higher heritabilities in our case. This could be explained by the difference of k in the k-fold cross validation between both studies and the fact that we did not divided our TS by population. A future improvement could be to adopt the same analytical design as explained in the cited publication where authors focused on predictive ability into single populations.

QTL detected in cross validation has been compared to QTL initially detected on the overall three populations and the QTL detection suffered some inconsistencies (data not shown). This can be associated with size of population used for QTL detection where all recombination are not sampled in the ES. Same results have been already observed by Melchinger et al. (1998) where only limited number of QTL have been commonly detected from two different samples of progenies from a same pedigree. The authors concluded that, QTL effects estimated from an independent sample can deviate when compared with effect estimated on another sample. This results inevitably in an overly optimistic assessment of the efficiency of MAS.

Whole Genome based prediction of unobserved genotypes.

The cross validation of GBLUP estimated breeding values showed higher consistency compared to QTL based estimated breeding values. This deviation can be explained by a non-consistent QTL effect estimation when QTLs have been detected on an independent set of progenies. The whole genome marker information held by the GBLUP model can also explain a greater part of the variance captured by the prediction model. The values obtained by the joint population model are in accordance with heritabilities of the traits (higher predictive ability for GM compared to GY). The predictive ability we found are consistent or slightly higher compared to the one found in other independent studies (Lian et al. 2014; Zhao et al. 2012). The cross validation showed reliable prediction accuracy across the different iterations of cross-validation. The joint-population and the within population scenarios of genomic selection showed the same predictive ability values (Table 4).

The structure of the breeding population does not impact the predictive accuracy. The main difference was observed when the joint-population scenario was compared with the across population scenario. The predictive ability dropped when predicting a population with a training set formed by other populations even if sharing a common parent (from 0.58 to 0.39 for GY and from 0.73 to 0.53 for GM). The close relationship between estimation set and test set has been already documented as potential threat to increase predictive ability (Windhausen et al. 2013). For the across population scenario, we just get an averaged predictive ability across the three populations and next step could define if the three populations behave the same way. Lehermeier et al. (2014) showed that predictive abilities similar to or higher than those within biparental families could be achieved by combining several half-sib families in the estimation set. In the same study, authors showed a large variance for predictive ability using the across population scenario, according to the population used as test set.

Conclusion

We demonstrated in this experiment the difference for predictive ability between whole genome prediction and QTL based prediction for two major traits for maize breeding. The lack of accuracy for QTL based prediction translated major issues that have been faced since many years now to efficiently play with QTL determinant of polygenic and complex trait. There is no predictive ability gain in combining several populations together in the training set compare to within biparental prediction only. In parallel, we showed the drop in predictive ability when predicting progenies from a single population without any representative in the training set, even if population were connected with a common parental line. This also calls the projection of genetic information from one population to another into question. Our results are important for defining future experimental design in whole genome prediction as they provide guidance to define the best genetic structure to be used for model training.

Acknowledgments

We wish to thank the Italian and Spanish Syngenta field trialing team for help in trial execution and phenotypic data collection.This study was supported by Syngenta Italy.

References

Abera Desta, Z,, Ortiz, R. (2014) Genomic selection: genome-wide prediction in plant improvement. Trend in Pant Science, 19: 592-601

Bauer E, Falque M, Walter H, Bauland C, Camisan C, et al. (2013) Intraspecific variation of recombination rate in maize. Genome Biol. 14: R103.

Bekavac G, Purar B, Jockovic D (2008) Relationships between line per se and testcross performance for agronomic traits in two broad-based populations of maize. Euphytica 162:363–369

Bernardo R, (1996) Best linear unbiased prediction of maize single-cross performance. Crop Sci. 36: 50–56

Bernardo R, Jianming Y. (2007). Prospects for Genomewide Selection for

Quantitative Traits in Maize. Crop Science Vol. 47 no. 3: 1082-1090

Bouchez A, Hospital F, Causse M et al. (2002) Marker-Assisted Introgression of favorable alleles at Quantitative Trait Loci between maize elite lines. Genetics 162:1945-1959

Charcosset A, Mangin B, Moreau L, Combes L, Jourjon M-F, Gallais A (2001) Heterosis in maize investigated using connected RIL populations. Quantitative genetics and breeding methods: the way ahead. Les colloques 96, INRA Editions, Paris, France

Charmet G, Robert N, Perretant MR, Gay G, Sourdille P, Groos C, Bernard S, Bernard M (1999). Marker-assisted recurrent selection for cumulating additive and interactive QTLs in recombinant inbred lines. Theoretical and Applied Genetics Volume 99, Issue 7, pp 1143-1148

Churchill, G.A., Doerge, R.W. (1994) Empirical threshold values for quantitative trait mapping. Genetics 138:963-71

Cooper M, van Eeuwijk FA, Hammer GL, Podlich DW, Messina C (2009). Modeling QTL for complex traits: detection and context for plant breeding, Current Opinion in Plant Biology, Volume 12, Issue 2: 231-240

De Roos AP, Hayes WBJ, Goddard ME (2009) Reliability of genomic predictions across multiple populations. Genetics 183: 1545–1553

Duvick D (1999) Heterosis: feeding people and protecting natural resources, pp. 19–29 in The Genetics and Exploitation of Heterosis in Crops, edited by J. Coors, and S. Pandey. CSSA, Madison, WI.

Foiada F; Westermeier P; Kessel B; Ouzunova M; Wimmer V; Mayerhofer W; Presterl T; Dilger M; Kreps R; Eder J; Schön CC (2015) Improving resistance to the European corn borer: a comprehensive study in elite maize using QTL mapping and genome‑wide prediction. Theor Appl Genet. 128:875–891

Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES 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:e28334.

Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389-2397

Haldane JBS (1919) The combination of linkage values, and the calculation of distance between the loci of linked factors. J Genet 8:299-309

Hallauer AR, Miranda JB (1988) Quantitative genetics in maize breeding. Iowa State University Press, Ames

Hallauer A. (1990) Methods used in developing maize inbreds. Maydica 35:1–16.

Hallauer AR, Carena MJ (2009) Maize breeding. In: Carena MJ (ed) Cereals. Springer, New York, pp 3–98

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009). Invited review: Genomic selection in dairy cattle: Progress and challenges. Journal of Dairy Science, Volume 92, Issue 3, March 2009, Page 1313

Heffner EL, Sorrells ME, Jannink JL (2009). Genomic Selection for Crop Improvement. Crop Science Vol. 49:1-12.

Herrera M, Conchello P, Juan T, Estopanan G, Herrera A, Arino A (2010) Fumonisins concentrations in maize as affected by physico-chemical, environmental and agronomical conditions. Maydica 55:121–126

Hill WG, Robertson A (1966) The effect of linkage on limits to artificial selection. Genet. Res. 8: 269–294.

Holm, S. (1979) A simple sequentially rejective multiple test procedure. Scand J Stat:65-70

Jannink, JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Brief. Funct. Genomics Proteomics 9: 166–177.

Jourjon, M-F., Jasson, S., Marcel, J., Ngom, B., Mangin, B. (2005) MCQTL: multi-allelic QTL mapping in multi-cross design. Bioinformatics 21:128-130

Lehermeier C, Krämer N, Bauer E, Bauland C, Camisan C, Campo L, Flament P, Melchinger AE, Menz M, Meyer N, Moreau L, oreno-González J, Ouzunova M, Pausch H, Ranc N, Schipprack W, Schönleben M, Walter H, Charcosset A, Schön CC (2014). Usefulness of Multiparental Populations of Maize (Zea mays L.) for Genome-Based Prediction. Genetics, Vol. 198, 3–16.

Lian L, Jacobson A, Zhong S, Bernardo R (2014). Genomewide Prediction Accuracy within 969 Maize Biparental Populations. Crop Science Vol. 54 no. 4: 1514-1522

Lippman ZB, Zamir D 2006 Heterosis: revisiting the magic. Trends Genet. 23: 60–66

Luan, T., Woolliams, J.A., Lien, S., Kent, M., Svendsen, M., Meuwissen, T.H.E. (2009) The accuracy of genomic selection in Norwegian red cattle assessed by cross-validation. Genetics 183:1119-1126

Maenhout S, De Baets B, Haesaert G (2010) Prediction of maize single-cross hybrid performance: support vector machine regression vs. best linear prediction. Theor. Appl. Genet. 120:415–427

Mangin B, Cathelin R, Delannoy D, Escalière B, Lambert S, Marcel J, Ngom B, Jourjon M-F, Rahmani A, Jasson S (2010) MCQTL: a reference manual. Rapport Ubia Toulouse N° 2010/1 - February 2010. Département de Mathématiques et Informatique Appliquées. Institut National de la Recherche Agronomique (INRA), France

Martin RA, Johnston HW (1982) Effects and control of Fusarium diseases of cereal grains in the Atlantic province. Can J Plant Pathol 4:210–216

Mather DE, Kannenberg LW (1989) Correlations between grain yield and percentage grain moisture at harvest in Ontario hybrid corn trials. Can J

Plant Sci 69:223–225

Massman, JM, Gordillo A, Lorenzana RE, and Bernardo R (2013) Genome wide predictions from maize single-cross data. Theor. Appl. Genet. 126: 13–22

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H et al. (2009) Genetic properties of the maize nested association mapping population. Science 325: 737–740.

Melchinger AE, Gumber RK (1998) Overview of heterosis and heterotic groups in agronomic crops, pp. 29–44 in Concepts and Breeding of Heterosis in Crop Plants, edited by Lamkey KR and Staub JE. CSSA, Madison, WI.

Melchinger AE, Friedrich Utz H, Schön CC (1998). Quantitative Trait Locus (QTL) Mapping Using Different Testers and Independent Population Samples in Maize Reveals Low Power of QTL Detection and Large Bias in Estimates of QTL Effects. Genetics May 1998 149:383-403

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Moreau L, Charcosset A, Hospital F, Gallais G (1998). Marker-Assisted Selection Efficiency in Populations of Finite Size. Genetics March 1998 148:1353-1365

Paterson A, Lander ES, Hewitt JD et al. (1988) Resolution of quantitative traits into mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721-726

Peleman JD, van der Voort JR (2003) Breeding by Design. Trends in Plant Science 8 (7):330-334

R Development Core Team (2013) R: A Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Reif JC, Zhao Y, Würschum T, Gowda M, Hahn V (2013) Genomic prediction of sunflower hybrid performance. Plant Breed. 132: 107–114

Röber, F.K., Gordillo, G.A., Geiger, H.H. (2005) In vivo haploid induction in maize – performance of new inducers and significance of doubled haploid lines in hybrid breeding. Maydica 50:275-283

Robertson-Hoyt LA, Betran J, Payne GA, White DG, Isakeit T, Maragos CM, Molnar TL, Holland JB (2007b) Relationships among resistances to

Fusarium and Aspergillus ear rots and contamination by fumonisin and aflatoxin in maize. Phytopathology 97:311–317

Rogers JS (1972) Measures of genetic similarity and genetic distances. Studies in Genetics, Univ Texas Publ 7213:145-153

Sala RG, Andrade FH, Camadro EL, Cerono JC (2006) Quantitative trait loci for grain moisture at harvest and field grain drying rate in maize (Zea mays, L.). Theor Appl Genet 112:462–471

Sala RG, Andrade FN, Cerono JC (2012) Quantitative trait loci associated with grain moisture at harvest for line per se and testcross performance in maize: a meta-analysis- Euphytica (2012) 185:429-440

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326: 1112–1115.

Shull, G. H., 1908 The composition of a field of maize. J. Hered. os-4: 296–301.

Silva Dias JC (2010) Impact of improved vegetable cultivars in overcoming food insecurity. Euphytica 176: 125–136

Stuber CW, Edwards MD, Wendel JF (1987) Molecular marker-facilitated investigations of quantitative trait loci in maize. II Factors influencing yield and its component traits. Crop Sci 27:639-648

Tardieu F, Tuberosa R (2010). Dissection and modelling of abiotic stress tolerance in plants. Current Opinion in Plant Biology, Volume 13, Issue 2: 206-212,

Technow F, Schrag TA, Schipprack W, Bauer E, Simianer H, Melchinger AE (2014) Genome Properties and Prospects of Genomic Prediction of Hybrid Performance in a Breeding Program of Maize. Genetics, 197:1343–1355

Tuberosa R, Salvi S (2006) Genomics-based approaches to improve drought tolerance of crops. Trends Plant Sci 11:405-412

Van Ooijen JW (2006) JoinMap ® 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, Netherlands

Wedzony MB, Forster I, Zur E, Golemiec M, Szechynska-Hebda et al. (2009) Progress in doubled haploid technology in higher plants, pp. 1–33 in Advances in Haploid Production in Higher Plants, edited by A. Touraev, B.

Forster, and S. Jain. Springer-Verlag, Dordrecht, The Netherlands.

Wimmer V, Albrecht T, Auinger H-J, Schön C-C (2012) synbreed: a framework for the analysis of genomic prediction data using R. Bioinformatics 28:2086-2087

Windhausen VS, Atlin GN, Hickey JM, Crossa J, Jannink JL, Sorrells ME, Raman B, Cairns JE, Tarekegne A, Semagn K, Beyene Y, Grudloyma P, Technow F, Riedelsheimer C, Melchinger AE (2013) Genomic Prediction in Maize Breeding Populations with Genotyping-by-Sequencing. G3 November 6, 2013 3:1903-1926

Xiang K, Reid LM, Zhang Z-M, Zhu X-Y, Pan G-T (2012) Characterization of correlation between grain moisture and ear rot resistance in maize by QTL meta-analysis. Euphytica (2012) 183:185–195

Zhao Y, Zeng J, Fernando R, Reif JC (2013) Genomic prediction of hybrid wheat performance. Crop Sci. 53: 802–810

Zhao Y, Gowda M, Liu W, Würschum T, Maurer HP, Longin FH, Ranc N, Reif JC (2012). Accuracy of genomic selection in European maize elite breeding populations. Theoretical and Applied Genetics Volume 124, Issue 4: 769-776

Figure legends

Figure 1. Box plots of Grain Yield (GY) and Grain Moisture (GM) data in the eight testing locations for the whole data set comprising the three populations in testing

Figure 2. Consensus map built using the 1164 SNPs segregating in at least one population description: Chromosome dimension, number of markers mapped in each chromosome and density of markers detected across the genome

Figure 3. Realized genetic relationship matrix between the DH lines of the three populations in testing

Supplemental material legends

Supplemental material 1. Coefficient of variation (cv) for Grain Yield for the eight yield testing locations. Data of the commercial hybrids used as reference checks and repeated several times in each location were used to calculate the cv value

Supplemental material 2. ANOVA table for Grain Yield and grain Moisture in the data set used. The full data set of the three populations in testing was used to make the calculation

Supplemental material 3. Heritability (h^2) for Grain Yield and grain Moisture in the data set used. The full data set of the three populations in testing was used to make the calculation

Supplemental material 4. Box plots of Grain Yield (GY) and Grain Moisture (GM) BLUEs data for each of the three populations in testing

Supplemental material 5. Roger genetic similarity between the four lines (L1, L2, L3 and L4) used as parents for the multi-parental DH population in testing and the line used as tester in testcross production L5.

Table 1. Phenotypic data description for each of the three populations in testing for the two collected traits (Grain Yield and Grain Moisture)

Grain yield (q ha^{-1})						
Population	**Minimun**	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Maximum**
Pop-1	50.8	125.0	135.4	134.8	145.6	189.5
Pop-2	63.0	122.8	134.5	133.2	144.7	184.0
Pop-3	57.2	119.3	131.7	130.8	142.6	192.2
Grain moisture (%)						
Population	**Minimun**	**1st Quartile**	**Median**	**Mean**	**3rd Quartile**	**Maximum**
Pop-1	9.6	14.9	19.0	18.9	22.6	32.5
Pop-2	10.0	13.3	17.9	18.1	22.2	28.2
Pop-3	10.0	13.2	18.3	18.2	22.1	29.7

Table 2. Chromosome (Chr.) position (Pos.), LOD score at the QTL position, proportion of variance explained (R^2) and additive effects of QTL alleles derived from parents L1, L2, L3 and L4 detected in the joint analysis across the three populations evaluated as testcross in 2014 for the traits Grain Yield at standard moisture (GY) and Grain Moisture at harvest (GM) traits. In bold is indicated the total R^2 for QTL detected simoultaneously

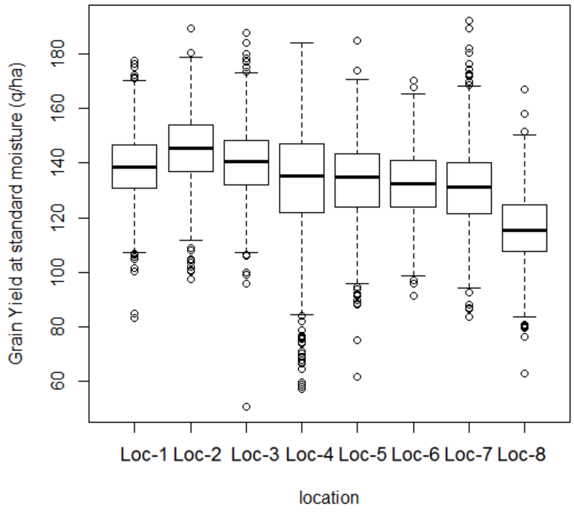
Chromosome	**QTL position (cM)**	**LOD**	**R^2**	**Additive** Effect			
				L1	**L2**	**L3**	**L4**
Grain Moisture (GM)			**0.41**				
1	0	3.21	0.02	-0.09	0.07	-0.03	0.05
1	140	6.56	0.04	0.03	0.11	-0.06	-0.07
1	221	10.61	0.06	-0.16	0.04	0.03	0.09
2	63	11.34	0.06	-0.03	0.06	0.1	-0.13
2	144	6.32	0.04	0.01	-0.04	0.12	-0.09
3	70	20.38	0.10	0	0.19	0.01	0.18
4	69	16.02	0.08	0.09	0.06	0.2	0.17
4	103	22.85	0.12	0.07	-0.2	-0.06	0.19
5	35	11.64	0.06	-0.09	0.02	0.24	-0.16
5	95	14.43	0.08	0.07	0.07	0	-0.14
6	53	7.52	0.04	-0.11	-0.02	0.05	0.08
7	67	8.03	0.04	0.05	0.06	-0.01	-0.1
8	105	9.54	0.05	0.03	0.1	0.04	0.11
9	85	4.62	0.03	0.02	0.01	0.09	0.08
Sum of additive effects				**-0.11**	**0.53**	**0.72**	**0.26**
Grain Yield (GY)			**0.37**				
1	174	29.12	0.14	-0.51	-0.21	-1.22	1.93
2	20	7.17	0.04	0.26	-0.08	0.77	-0.95
3	71	7.10	0.04	0.2	-0.26	1.01	-0.94
3	178	6.17	0.03	0.43	-1	-0.13	0.69
5	12	10.37	0.05	-1.03	0.53	-0.53	1.02
5	94	5.36	0.03	0.06	-0.26	-0.61	0.8
6	80	4.31	0.02	0.04	-0.7	0.01	0.65
8	88	8.64	0.05	0.54	-1.54	0.41	0.6
10	93	3.26	0.02	0.35	-0.74	0.66	0.43
Sum of additive effects				**0.34**	**-4.26**	**0.37**	**4.23**

Table 3. Mean predictive abilities obtained from cross-validation of the QTL-based and GBLUP models for joint-population prediction for the GY and GM traits

Prediction Model	**GY**	**GM**
QTL	0.15 ± 0.07	0.30 ±0.12
GBLUP	0.5800 ± 0.0062	0.7335 ± 0.0024

Table 4 GBLUP predictive abilities from within and across-population prediction

Analysis type	**GY**	**GM**
Within-Population GBLUP	0.5759 ± 0.0055	0.7318 ± 0.0038
Across.Population GBLUP	0.3881 ± 0.0031	0.5306 ± 0.0181



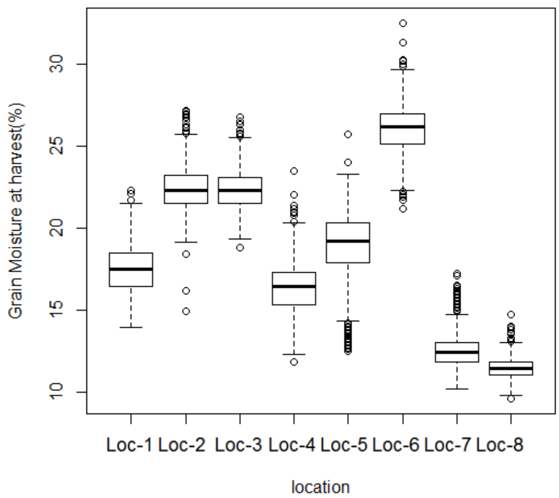


Figure 1. Box plots of Grain Yield (GY) and Grain Moisture (GM) data in the eight testing locations for the whole data set comprising the three populations in testing

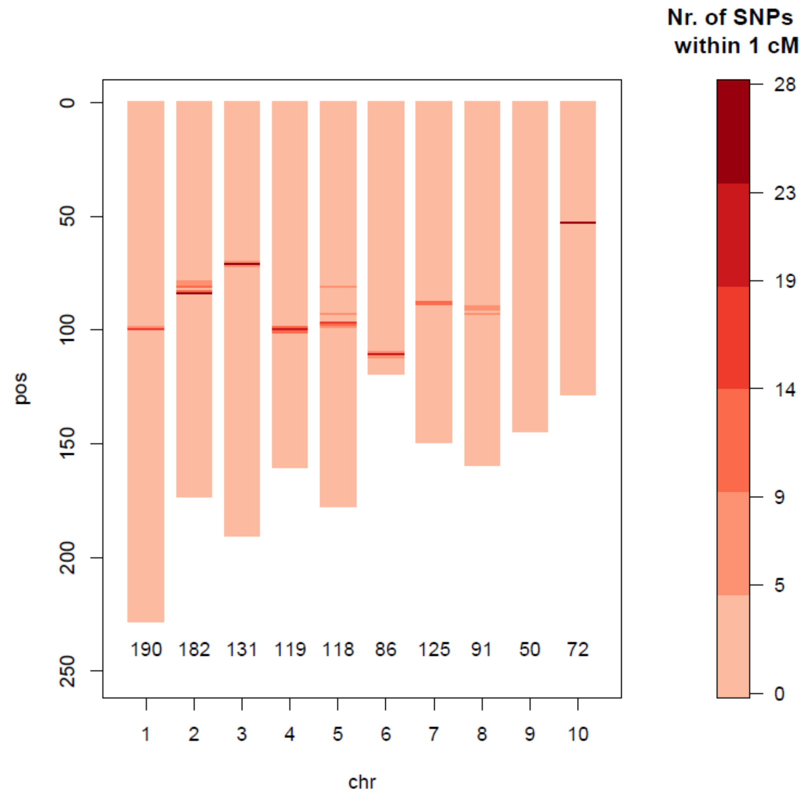


Figure 2. Consensus map built using the 1164 SNPs segregating in at least one population description: Chromosome dimension, number of markers mapped in each chromosome and density of markers detected across the genome

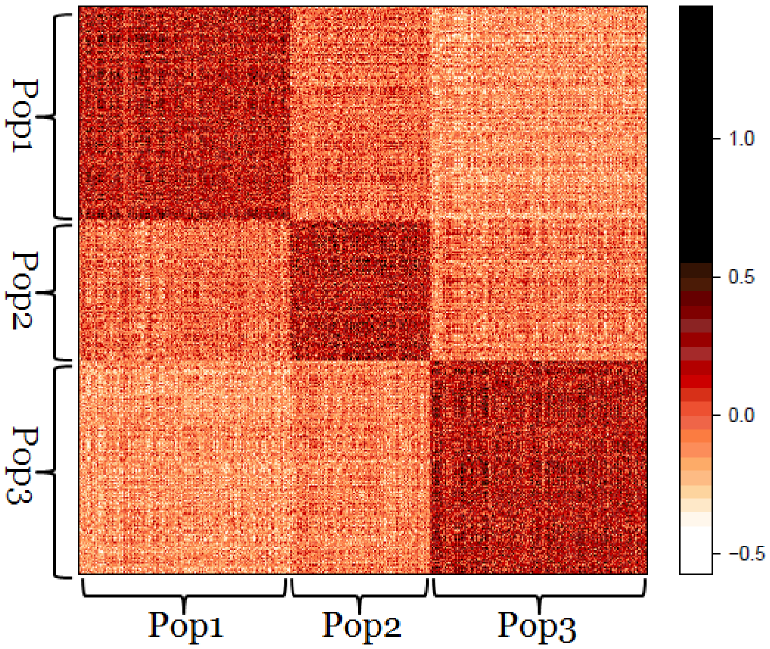


Figure 3. Realized genetic relationship matrix between the DH lines of the three populations in testing

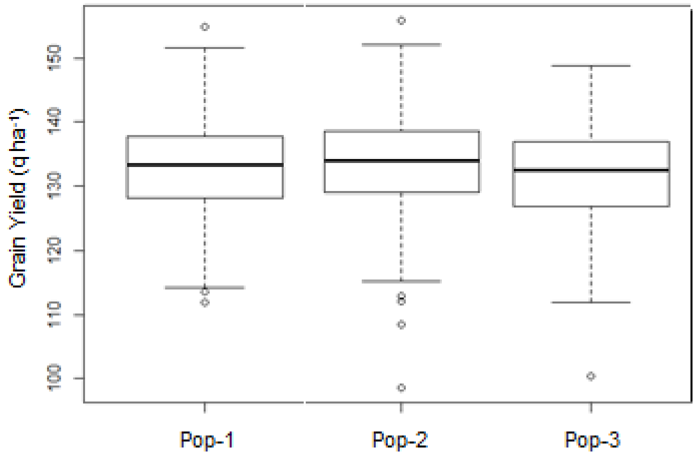
Supplemental material 1. Coefficient of variation (cv) for Grain Yield for the eight yield testing locations. Data of the commercial hybrids used as reference checks and repeated several times in each location were used to calculate the cv value

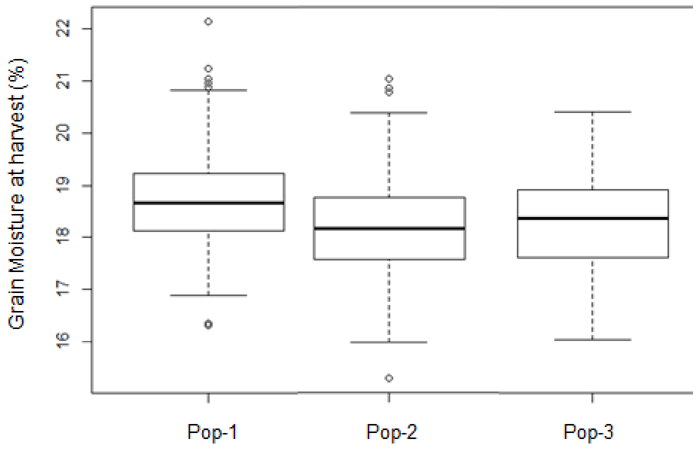
Location	**cv (%)**
Loc-1	7%
Loc-2	9%
Loc-3	9%
Loc-4	11%
Loc-5	7%
Loc-6	9%
Loc-7	12%
Loc-8	10%

Supplemental material 2. ANOVA table for Grain Yield and grain Moisture in the data set used. The full data set of the three populations in testing was used to make the calculation.

		Df	Sum Sq	Mean Sq	F value	Pr(>F)	Significance
GY	Genotype	949	456079	481	4.07	$< 2.2e^{-16}$	***
	location	7	520038	74291	629.35	$< 2.2e^{-16}$	***
	Genotype x Location	6605	1164901	176	1.4941	$2.93E^{-10}$	***
	Residuals	582	68701	118			
GM	Genotype	949	7664	8.1	8.83	$< 2.2e^{-16}$	***
	location	7	182212	26030.3	28488.94	$< 2.2e^{-16}$	***
	Genotype x Location	6605	9493	1.4	1.57	$1.94e^{-12}$	***
	Residuals	582	532	0.9			

Supplemental material 3. Heritability (h^2) for Grain Yield and grain Moisture in the data set used. The full data set of the three populations in testing was used to make the calculation	
Trait	**Heritability (h^2)**
Grain Yield (GY)	0.64
Grain Moisture (GM)	0.82





Supplemental material 4. Box plots of Grain Yield (GY) and Grain Moisture (GM) BLUEs data for each of the three populations in testing

Supplemental material 5. Roger genetic similarity between the four lines (L1, L2, L3 and L4) used as parents for the multi-parental DH population in testing and the line used as tester in testcross production L5.

Inbred	**L2**	**L3**	**L4**	**L5**
L1	0.72	0.82	0.59	0.57
L2		0.77	0.61	0.56
L3			0.60	0.59
L4				0.53