ASSOCIATION MAPPING OF AGRONOMIC TRAITS OF DRY BEANS USING BREEDING POPULATIONS

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Chiti Agarwal

In Partial Fulfillment
for the Degree of
MASTER OF SCIENCE

Major Department:
Plant Sciences

May 2014

Fargo, North Dakota

North Dakota State University

Graduate School

Title

ASSOCIATION MAPPING OF AGRONOMIC TRAITS OF DRY BEANS USING BREEDING POPULATIONS

By

Chiti Agarwal

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. Juan M. Osorno
Chair

Dr. Phil McClean

Dr. Kevin McPhee

Dr. Eugene Berry

Approved:

5/5/14	Richard D. Horsley
Date	Department Chair

ABSTRACT

Genome wide association mapping (GWAS) is an effective method to fine-map QTL because of its higher mapping resolution. In order to evaluate the possibility of using breeding populations for GWAS, analysis were conducted using AYTs (Advanced Yield Trials) and PYTs (Preliminary Yield Trials) from the NDSU dry bean breeding program, grown in 2012 at four locations in North Dakota using a 6k SNP chip. Genomic regions were evaluated separately for AYT, PYT, AYT+PYT, and races Mesoamerica and Durango. Overall, 13, 11, 9, and 9 significant markers were found for seed yield, maturity, 100-seed weight, and plant height respectively. Two candidate genes for seed yield and four candidate genes for days to maturity were identified. These markers are highly diagnostic within and among NDSU bean breeding populations and therefore, they could be directly used in Marker assisted selection to develop improved bean varieties while maintaining commercially desired phenotypic characteristics of beans.

ACKNOWLEDGMENTS

I would like to thank Dr. Juan M. Osorno for his guidance, insight, assistance, and support. I could not have finished this without his help. I would also like to express my greatest gratitude to my committee members: Dr. Kevin McPhee, Dr. Phil McClean, and Dr. Eugene Berry, for their helpful comments. I would like to give special thanks to Dr. Sujan Mamidi for his guidance and support in my whole research.

I would like to give special thanks to my Mom, Dad and family as the pursuit of a Master's Degree would have never been attempted or completed without their support.

I will not forget the kindness help from the staff of the Dry Bean Breeding Program at NDSU, Angela M. Linares-Ramírez, Kiran Ghising, José E. Vazquez, Jody Vander Wal as well as Rian Lee and Samira Mafi Moghaddam from the bean molecular genetics group.

I am grateful to have had the classmates I did. Completion of my degree was easier, interesting and eventful having you guys and girls as classmates.

I have great appreciation for Northarvest Bean Growers Association, the BeanCAP project funded by USDA-NIFA, and the Department of Plant Sciences at North Dakota State University. Without their support, this opportunity would not have been available.

Finally, thanks to all my friends who supported me in all possible ways during this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Association mapping is a method for mapping quantitative trait loci (QTL) that takes advantage of linkage disequilibrium (LD) to find associations between phenotype and genotype. In plants, association mapping exploits historical evolutionary events and helps in studying complex traits of both economic as well as agronomic importance. Dry bean is an important cash crop with high nutritional value and is produced over more than 1.3 million ha in United States (Singh *et al*., 2005) contributing 1.8$ billion to the U.S. economy (McClean *et al*., 2004). Dry beans exhibit extensive extent of linkage disequilibrium, due to its narrow genetic base caused mostly to breeding and population bottleneck during domestication and selection for cultivars within each market class (Rossi *et al*., 2009). Genome wide association mapping (GWAS) has been conducted in many other crops such as rice (*Oryza sativa* L.), maize (*Zea mays* L.) , barley (*Hordeum vulgare* L.) using diversity panels, biparental populations, variety trials etc. to find markers associated with seed yield, days to flowering, disease resistance, etc (Ersoz *et al*., 2008, Jianming *et al.*, 2006). GWAS in beans using the 6K SNP chip (Cregan and Qijian, 2012) and lines from breeding programs could also be used as mapping populations and can, therefore, help in finding significant markers associated with agronomic traits and potential candidate genes that control these traits. Identification of highly diagnostic markers within and among breeding populations could provide an opportunity to develop improved bean varieties in a more efficient way when incorporated into the breeding program. This will help the breeders in enhancement of genetic diversity while maintaining commercially desired phenotypic characteristics of beans.

The main objective of this research is to evaluate the efficiency of GWAS to discover marker-trait associations for seed yield, 100-seed weight, days to maturity, and plant height in dry beans using the breeding populations normally used within the NDSU dry bean breeding

program; to identify genomic regions involved in controlling agronomic traits of economic importance; to identify possible candidate genes with known functions which fall within the significant QTL regions.

LITERATURE REVIEW

Importance of dry beans

Dry bean (*Phaseolus vulgaris* L.) is the most important grain legume for direct human consumption (Kelly, 2010). It is consumed in many countries of the world and mostly preferred in Latin America, the Caribbean, and Eastern Africa. The top producers of dry bean in the world are Latin America, with Brazil, Mexico and the U.S. as major producers and second largest is Africa with Uganda, Kenya, Rwanda, Burundi, Tanzania, and Congo as a major producer (CGIAR, 2013). *P. vulgaris* L. is an edible legume which provides 30% of the total daily calories to the world's population (Kalavacharla *et al*., 2011). This fiber rich diet staple, is also rich in protein, carbohydrates, vitamins, minerals, and is free of saturated fat and trans-fat, and cholesterol (Azarpazhooh and Boye, 2012).

Economic importance of crop

Commercial production of dry beans in USA started in early 20th century. Today, dry beans are grown in more than 30 states of U.S. The top dry bean producing states are North Dakota (38%), Michigan (14%), Nebraska (11%), Minnesota (10%), Idaho (7%), and Colorado (5%). In 2013, dry beans were planted in U.S. on 0.5 million hectares with an average seed yield of 2091 kg ha^{-1} with a total production of 1.2 million MT. It was first commercially grown in North Dakota in 1962 and later during the 1990s, North Dakota has become the leading producer of dry beans in the U.S. In North Dakota, estimated total production was 0.3 million MT harvested from 250,905 ha where pinto (242,073.17 MT) is the most predominant market class followed by navy (65,992.24 MT), and black (27,687 MT) (USDA-NASS, 2013). In Minnesota, the production of dry beans was 0.1 million MT, harvested from 68,796 ha, where production of

kidney beans (45,722 MT) leads, followed by navy beans (35,053.61 MT), black (24,689.65 MT), and pinto (9,347.63 MT) (USDA-NASS, 2013).

Taxonomy of dry beans

Dry bean is a member of the monophyletic genus *Phaseolus* (Delgado-Salinas *et al*. 1999; 2006). It is a highly variable annual plant native to many cultures around the world. It belongs to subtribe Phaseolinae. It is a member of the tribe Phaseoleae, subfamily Papilionoideae, and the family Fabaceae (Leguminosae). Other genera in the Papilionoidae subfamily include *Glycine*, *Vigna*, *Pisum*, *Cicer*, and *Medicago*. There are about 50 wild species in the *Phaseolus* genus, out of which five species have been domesticated, namely *P. vulgaris* or common bean; *P. polyanthus* or year-long bean; *P. coccineus* or scarlet runner bean; *P. acutifolius* or tepary bean; and *P. lunatus* or lima bean (Singh *et al*., 2005). These species represent a wide range of life histories (annual to perennial), growth habits (bush to climbing), reproductive systems, and adaptations (from cool to warm and dry to wet). Common bean is predominantly self-pollinated and it is grown in temperate conditions.

Domestication and organization of the genetic diversity

Recent evidence suggest that dry bean's center of origin is located in Mesoamerica (Bitocchi *et al*., 2011). Depending upon the several morphological traits, and the study based on molecular markers, the wild common bean is commonly distinguished into two gene pools (Beebe *et al*., 2001, Gepts, 1991; Gepts, 1998; Singh *et. al*., 1991; Bitocchi *et. al*., 2012). The Middle American gene pool extends from Mexico through Central America while Andean gene pool is found in Colombia, Ecuador, Peru, Chile, Bolivia, and Argentina. Both gene pools diverged at about 110,000 years ago and underwent through a genetic diversity bottleneck. After divergence, the Andean and Middle American bottlenecks started at 103,000 years ago and

ended 62,000 years ago. Later, due to domestication bottleneck, there was a reduction in diversity and increase in population structure (Schmutz *et al*., 2011). In addition, a third gene pool of wild bean populations were identified between the region of Peru, Ecuador, and Colombia and were characterized by the presence of a specific seed storage protein known as Phaseolin type I, and allozymes (Singh *et al*., 1991a; 1991b). Since it shares alleles from both gene pools it is known as introgression between Andean and Mesoamerica. In dry bean populations, many changes are observed by looking to morphological traits (changes in size for leaves, pods, and seeds, loss of pod dehiscence, increase in permeability of seed coat, decrease in anti-nutritional factors (Smartt, 1988; Gepts, 1991), and molecular markers (Becerra Velasquez and Gepts, 1994; Tohme *et al*. 1996; Bitocchi *et. al*., 2012), environmental factors in agronomic characteristics and also in terms of partial reproductive isolation i.e. the fatality of F_1 generation in some crosses, in both wild and domesticated species (Gepts, 1998).

Genotypes from the Andean pool usually show smaller bracteoles, and usually only one pod-bearing node compared to two nodes of the Middle American genotypes (Gepts and Debouck, 1991). Other changes in the morphological features among two gene pools include the length of the internode and flowering time, where Andean genotypes flower earlier than Mesoamerican genotypes due to its determinate growth habit (Gepts and Bliss 1986; Singh *et al*. 1991b).

In Mesoamerican genotypes, the presence of S-type phaseolin with no T-type contrasted to Andean genotypes by the presence of mostly T-type phaseolin with C, S, H, and A types, but in low frequencies (Gepts, 1998). Further, Kami *et al*. (1995) confirmed the close relative of type I phaseolin in accessions from northern Peru-Ecuador to the other phaseolin in wild *P. vulgaris*, and thus suggesting the dispersal of wild bean into north (Colombia, Central America, and

Mexico) and south (southern Peru, Bolivia, and Argentina) that led to two gene pools i.e. Mesoamerican and Andean. However, several studies on the phaseolin types, allozyme alleles, and molecular markers have revealed Mesoamerican gene pool to be the origin of common bean with higher diversity (Bitocchi *et al*. 2012). Additionally, data on the amplified fragment length polymorphisms (AFLP), and simple sequence repeats (SSRs) of wild and domesticated *P. vulgaris* have re-proposed the origin of the common bean within the Mesoamerican gene pool (Kwak and Gepts, 2009; Rossi *et al*., 2009). Similarly, Bitocchi *et al*. (2012) investigated the nucleotide diversity at five gene loci of a large sample that represented the entire geographical distribution of the wild form of *Phaselous* species, and confirmed Mesoamerica as the origin of the common bean. The results showed that in Andean gene pool loss of genetic diversity during and after domestication was much more as compared to Mesoamerican gene pool.

The Mesoamerican gene pool is divided into three races: Durango in the central highlands of Mexico, Jalisco in coastal Mexico, and Mesoamerica in lowland tropical Central America. Race Guatemala has been proposed as a fourth race located in the central region of Central America, and consists mostly of medium sized black-seeded climbing genotypes (Beebe *et. al.* 2001). Pinto, great northern, small red, and pink bean market classes belong to race Durango, while Mesoamerica race includes navy and black beans, with market classes Flor de Mayo and Flor de Junio (Mexico) in race Jalisco. The Andean gene pool is also subdivided into three races: Nueva Granada (Colombia/Ecuador), Peru (Peruvian highlands), and Chile (northern Chile and Argentina). Market classes light red kidney, dark red kidney, white kidney, and cranberry beans represent race Nueva Granada. Mayocoba and Canario also known as "yellow beans", represents Peruvian race, and race Chile includes the vine cranberry beans and bean types distinctive to Chile (Coscorron and Tortola) (Singh *et al*., 1991 and Kelly, 2010).

Several morpho-agronomic traits and ecological criteria have distinguished domesticated beans from their wild ancestors (Beebe, *et al*. 2001). Some of the differences include reduced plant height, shorter internodes and fewer nodes, non-dehiscent pods, and higher germination rate in domesticated types. For example, race Nueva Granada represents medium to large seeded accessions mostly with bush type growth habits, while Race Peru consists mostly of Andean climbing beans adapted to highland environments above 2,000 masl. Race Chile is characterized by prostrate type 3 growth habit, medium-sized, and rounded to oval seed (Beebe, *et al*. 2001).

Dissemination of the domesticated lines

From the centers of domestication, wild common bean is widely distributed throughout the highlands of Mesoamerica and South America to wet lowlands in Central America and high altitudes in the Andes (Chacon *et al*. 2005). The distribution of domesticated beans throughout North and South America was possible by human activities such as agriculture, urbanization, and deforestation (Debouck *et al*. 1993). This distribution comprises of three ranges of expansion events i.e. two from Mesoamerica to northern South America and one from the northern Andes to Central America. The distribution of Mesoamerican races of common bean to northern and eastern South America was made possible as a result of migration of Native Americans from north. The pathway of dissemination of common bean outside the American continent started after 1492. Common bean was introduced from America to Africa and then to Europe by the Spaniards and Portuguese (Maras and Sustar-Vozlic, 2013).

Dry bean development stages

Dry bean is recognized as a short day, self-pollinated crop. Slightly cool environment is favorable for bean growth and development. The growth habit of dry bean is based on growth habits during flowering and can be characterized as determinate or indeterminate. A determinate

plant exhibits a bush growth habit and an indeterminate/climbing plant is characterized by a vining/trailing pattern. When flowering begins in determinate plants, upward growth of stem stops, while in plants with indeterminate habit, stems continue to grow throughout flowering. (Kelly, 2010).

There are four different growth types. Growth type I is determinate bush and growth type II is indeterminate and upright short vine. Both have narrow plant profile with three to four stems. Growth type III is indeterminate and has prostrate vines. Growth habit type IV is indeterminate and has strong climbing tendencies. (Singh, 1982).

Agronomic traits of dry beans

There are several agronomic traits that breeders evaluate while developing cultivars in order to obtain maximum performance of dry bean genotypes. Some of these traits are the days of maturity, 100-seed weight, seed shape, plant height, days to flowering, disease resistance, tolerance to abiotic stress, and seed yield, among others.

Breeders look for quantitative traits considering economic needs and genetic limitations. Traits which are controlled by many different genes with low heritability are difficult to manipulate by cross breeding. Identification of markers linked or associated to the QTLs of these traits would benefit breeders by facilitating selection, especially at early generations (Chengsong *et al*., 2008).

The NDSU bean breeding program use a modified pedigree method for cultivar development. For this, F_1 seeds are generated from the crosses are produced in the greenhouse in Fargo, ND and are evaluated in New Zealand during winter based on their appearance and vigor. F_2 are selected and evaluated in again in North Dakota and its selected progenies (individual plants) are grown as F_3 rows in Puerto Rico during winter. Selected F_3 progenies are grown as F_4

rows back in North Dakota. From F_4, 3-4 uniform plants are selected based on vigor and appearance and bulked. F_5 plants which are then grown in Puerto Rico during winter and progenies of selected rows are bulked to test high value of quantitative traits and thus enter into preliminary yield trials (PYT) for yield testing and overall agronomic performance. These breeding lines are being tested in multiple locations over multiple years in North Dakota and Minnesota. Promising superior breeding lines enters advanced yield trials (AYT) if performed well. Disease screening is also made in the greenhouse for breeding lines included in both PYT and AYT. In the same way, canning quality is evaluated for the advanced lines. Best genotypes are then moved into variety trials to make final decisions about commercial release as cultivars. The entire breeding pipeline takes approximately 8-9 years to be completed from crossing to cultivar release.

Molecular characterization of *Phaseolus* species

Common bean is a diploid species with a chromosome number of 2n =2x =22, and an estimated genome size of 587 million base pairs (Mb) per haploid genome (Schmutz *et al.,* 2014).

Genetic diversity and variability of bean has been increased with the help of germplasm incorporation and inter-specific hybridizations. Initially, new varieties were selected and released directly from the landraces. Presently, many tools like fine mapping, GWAS etc. have been developed to aid in the genetic improvement of bean varieties and to study the relationships between genotype and phenotype, especially for complex traits. Several molecular markers have been utilized to develop linkage maps in common beans including Restriction Fragment Length Polymorphisms (RFLPs) (Adam-Blondon *et al*., 1994; Vallejos *et al*., 1992), Randomly Amplified Polymorphism DNAs (RAPDs) (Freyre *et al*., 1998), Amplified Fragment Length

Polymorphisms (AFLPs), Simple Sequence Repeats (SSRs), Resistance Gene Analogs (RGAs), Expressed Sequence Tags (ESTs) (McConnell *et al*., 2010) and InDel (Insertion-deletions) markers, that are developed by transposable elements, unequal crossover (Britten *et al*., 2003) and SNP markers identified by standard paired-end libraries of Illumina (Moghaddam *et al.,* 2014). Being highly abundant and informative these are widely used markers these days (Pacurar *et al.,* 2012).

With the help of these molecular techniques, many germplasm banks like CIAT work on the development of many PCR-based molecular markers, cDNA libraries, genomic libraries, and EST sequencing. More than 29,000 domesticated and 1,300 wild accessions of *P. vulgaris* are stored in this germplasm bank (Broughton *et al*., 2003).

The first molecular linkage map was developed by a backcross population from a cross between XR-235-1-1(Mesoamerican genotype) and 'ICA-Calima,' (Andean cultivar) (Vallejos *et al*. 1992). The map was made up of 224 RFLP markers, 9 isozyme markers, 9 seed protein markers, and a seed and flower color marker. The map covered 960 cM of the bean genome. This population is known as XC population. In this map, 11 linkage groups were established and were later used in an analysis for markers linked to disease resistance genes (Yu *et al*. 1998).

The second bean genetic map was constructed in an F_2 population from the cross of BAT93 (Mesoamerican) x Jalo EEP 558 (race Nueva Granada, Andean landrace) with total of 152 markers, where most markers were RFLPs and some RAPD. Both parents in the linkage map BAT93 × Jalo EEP 558 (BJ) showed high degree of phenotypic polymorphism, and disparity for many important traits, such as disease and insect reactions, seed nutritional quality and other morphological and agronomical traits. The total 143 markers that were assigned to 15 linkage groups covered 827 cM of the bean genome with 6.5 cM of an average interval between

the markers (Nodari *et al*., 1993a). The population is known as BJ. It segregates for resistance for many diseases like bean common mosaic virus or BCMV (*I* gene was mapped for resistance), common bacterial blight caused by pathogen *Xanthomonas axonopodis pv. phaseoli (Xap)* and *Xanthomonas. fuscans subsp*. *fuscans*, and fungal disease anthracnose caused by Anthracnose (*Colletotrichum lindemuthianum).*

A third map was developed in a backcross population derived from two European bean lines i.e. Ms8EO2 (resistant to anthracnose) x Corel (susceptible to anthracnose) (Adam-Blondon *et al*. 1994). The purpose of the map was to locate genes involved in plant defense and resistance against anthracnose, and used 51 RFLPs, 100 RAPDs and few SCARs and morphological markers. Thus developed markers were placed on 12 linkage groups that covered 567.5 cM of the genome.

Koinange *et al*. (1996) constructed a linkage map in RILs of a cross between Midas (wax snap bean cultivar) x G12873 (wild accession), where 83 RFLP were used and the population size were 65 recombinant inbred lines. This population is called MG population. This map was developed to discover loci associated with the domestication syndrome and domestication traits in common bean, which later was found the domestication QTL on 8 out of 13 linkage groups. Similarly, a reference map was developed using RIL population derived from BAT93 x Jalo EEP558 (BJ) population using microsatellites. Total of 106 markers were placed on 12 groups that covered total length of 606.8 cM of the genome. The purpose was to use this map for the development of many other molecular maps for common beans that focused on QTL and association studies (Grisi *et al*. 2007).

Recently, the genetic map was developed from the cross between PMB0225 x PHA1037 using 85 AFLP, 95 SSR, and 13 SNP markers, of which 101 were found dominant and 92 co-

dominants and formed 12 linkage groups. These groups were allotted on the basis of 55 previously mapped common SSR markers. The map covered total genetic distance of 822.1 cM, with an average of 68.5 cM/linkage group, ranging from 16.5 cM (LG 6) to 106.4 cM (LG 3). BIOAGRO/UFV and Embrapa bean research groups developed a map with a RIL population with 500 lines from the cross between Rudá (Mesoamerican) x AND 277 (Andean). Ruda is a high yielding Mesoamerican cultivar made from crosses between the cultivars (Carioca x Rio Tibagi) and AND 277 (Andean), developed by CIAT from multiple crosses [(Cargabello × (Pompadour Checa × Linea 17) × (Linea 17 × Red Cloud))]. Thus showing difference not only for diseases such as angular leaf spot, white mold, rust, and anthracnose, but also contrasts for many phenotypic traits like morphological and agronomical traits. They used 126 SSR and 677 SNP markers to estimate the genetic distance, which was 78.6% based on SSR and 71.3% based on SNPs. The estimated genetic distance of molecular marker maps ranges from 1259 cM to 1545.5 cM. The average distance between any two markers ranges from 3.0 cM to 7.23 cM (Sanglard *et al*., 2013).

Gepts in 1990 used data from isozymes, morphological and agronomic traits to determine the genetic diversity and to show the importance of genotype preservation in breeding programs. Molecular data of isozyme and phaseolin illustrated extensive hybridization between the Chile genotypes and genotypes found in Andean and Mesoamerican region. The research also suggested that the Chilean landraces can be split into two groups (Paredes and Gepts, 1995). Beebe *et al*. 2001 presented a study in which data of RAPD markers and DNA fingerprinting indicated that Guatemalan beans can be separated from southern Mexican beans. Skroch and Nienhuis (1995), using RAPD data showed that snap beans are intermediates between two gene pools rather than originated from the Andean gene pool. A similar study of diversity using

RFLPs, compared the wild and domesticated beans proving that they are from same region (Beccerra Velasquez and Gepts. 1994). Sonnante *et al*. 1994 indicated evolutionary changes like loss of genetic diversity during selection and domestication by using M13 (probe-enzyme combination) DNA fingerprinting.

Single nucleotide polymorphic (SNP) markers

Markers are easily selectable, highly heritable genetic "tags" that are linked to traits which are more difficult to select phenotypically. Thoday *et al*., (1961) suggested that markers can be used to map and characterize genes that control traits of interest. The development of molecular markers (isozymes and DNA-based markers) has created a potentially endless number of markers for analyzing genomes. SNP markers are currently known as valuable markers for genotyping because of their abundance, stability, and simplicity. In many species, these markers are distributed throughout the genome and are mostly these are the first choice for association mapping studies (Drenkard *et al*., 2000). SNPs represent most frequent polymorphisms not only in plant genome but also in human and animal genomes.

There are many different methods that are used for SNP genotyping in common bean. CAPS (Cleaved Amplified Polymorphic Sequences) and dCAPS (derived Cleaved Amplified Polymorphic Sequences) techniques that were employed to convert EST based polymorphisms into SNP markers. Cleavage nuclease CEL I technique by Galeano *et al*. was used to analyze and map SNP-based EST-derived markers. Single strand conformation polymorphism (SSCP) technology was also used to map EST-based markers. In DOR364 × G19833 mapping population, a total of 118 new marker loci was identified by using this high throughput technique. Luminex-100 was used to confirm SNP calls in DNA from 10 common bean genotypes. These techniques have some disadvantages like being unable to identify

polymorphism in contigs and amplicons with two or more SNPs, enzyme specific and expensive to use. KASPar technology is another technique used for SNPs detection that provide wider spectrum of genotyping. It is efficient and less time consuming as compared to all others (Cuenca *et al*., 2013).

SNP markers in common bean reflect dual domestication events and inter gene pool hybridization in both gene pools. They allowed the identification of two Andean and three Mesoamerican clusters corresponding to races. Due to greater polymorphism and race structure, Andean gene pool shows higher genetic diversity with SNPs than the Mesoamerican (Galeano *et al*., 2012; Cortes et al., 2011). The total number of SNPs in cultivated bean is estimated to be in the range of three to four million, based on the rate of 237 SNPs observed in 38.2 kbp of sequence in six diverse genotypes (Souza *et al*., 2009).

Association mapping

Association mapping is an effective method to map (QTL). It is a method of QTL mapping that links phenotypes to genotypes by using linkage disequilibrium (Yu *et al*., 2006). Association detection depends on genetic architecture (population structure), accurate phenotypic evaluations, and genotyping (Balding *et al.,* 2006). Association mapping is an alternative to bi-parental mapping. Association mapping is a preferred method over bi-parental mapping as it gives a higher mapping resolution and has greater allele number as compared to bi-parental mapping.

The major advantage of performing association mapping, by using a breeding pipeline instead of a diverse panel of genotypes is that the markers/QTL identified here are highly specific for the breeding program and can be used directly for further selection. Alternatively, using a diverse population reduces the chances of identifying markers/QTL that are relevant for

the breeding program because of the lack of representation of polymorphisms or recombinant events in the diverse panel. Also use of advanced lines targets multiple recombination events in association mapping approach and is limited to just a few and the probability of finding these increases compared to thousands of recombination events present at a low frequency in diverse panel (Podlich *et al*., 2004). Since bi-parental populations don't have to be developed, this mapping approach is not only less costly and less time consuming but also helps in marker assisted selections (MAS). Unlike association mapping, bi-parental mapping has fewer recombination events and is very expensive and time consuming (Zhu *et al*., 2008). Association mapping is also subject to some limitations; for example, in breeding populations the probability of identifying false positives may increase due to high population structure. In addition, high linkage disequilibrium can cause poor resolution.

Association mapping studies in plants was first reported in Oat (*Avena sativa* L.) and rice by Beer *et al*., 1997 and Virk *et al*., 1996 respectively. Oat association study used 64 oat varieties and showed that 13 QTLs are associated with restriction fragment length polymorphisms (RFLP) markers. In rice germplasm, six traits were predicted using RAPD markers. In maize, SNP markers were found to be associated with flowering time and plant height (Jafar *et al*., 2012). In barley (*Hordeum vulgare* L.), various traits such as seed yield and stability, heading date, flowering time, plant height, rachilla length, resistance to mildew and leaf rust were associated with many different types of molecular markers (Wang *et al*., 2011). In soybeans (Glycine max), association mapping was performed to discover molecular markers associated with iron deficiency chlorosis (Wang *et al*., 2008). QTLs for growth period traits in soybeans were identified by association mapping (Zuo *et al*., 2013). Later, association mapping was performed

using SNP markers in barley, wheat (*Triticum spp.*), potato (*Solanum tuberosum L.*) among other crops.

There are two types of association mapping: candidate association mapping and genome wide association mapping (GWAS) (Zhu *et al.,* 2008). GWAS is a practical approach for a bean domesticated population. This approach searches the whole genome for causal genetic variation. GWAS does not need any previous information on candidate genes and can test large number of markers for association with various complex traits. Due to the complex population structure present in beans and lack of information about candidate genes associated with agronomic traits, GWAS is the best approach that could be applied for association mapping to study agronomic traits. This large population structure and relatedness that exist in beans can identify false positives.

The first step of association mapping is selection of a germplasm, cultivars, or breeding lines etc. which has wide coverage of genetic diversity. Next, phenotypic characteristics of these natural populations are recorded. Then genotyping is done with available molecular markers. Next step involves quantification of the Linkage Disequilibrium (LD) and assessment of the population structure and kinship using the molecular marker data. Based on the information gained, a regression model is developed between phenotypic and genotypic data with appropriate statistical approach. This approach reveals "marker tags" that are present within close proximity of targeted traits of interest (Abdurakhmonov *et al*., 2008).

Yu *et al*., 2006 developed an approach to control relatedness and population structure in a regression model, known as the mixed model approach. In a mixed model approach, both fixed (SNP effect and population structure) and a random (kinship) effects are included, which makes it flexible to family based and population based samples. Mixed models represent a method of

choice that deals with unbalanced data across multiple trials. It shows reliable inference through the explicit modeling of correlations induced by genetic and environmental causes. Thus, this model is useful in genome wide association studies or association mapping to control the biased that may possibly be caused by the population structure and relatedness in other species.

Importance of phenotyping

Collection of high-quality phenotypic data is essential in association mapping. Newly discovered candidate genes in mapping studies can only be tested if we have existing robust and accurate phenotypic data, which is usually collected over years in multiple locations (Flint-Garcia *et al*., 2005). To increase the mapping power, it is necessary to consider efficient field designs with incomplete block design (e.g., α-lattice), appropriate statistical methods and QTL × environmental interaction (Eskridge, 2003).

OBJECTIVES

The main objective of this research is to evaluate the efficiency of GWAS in dry beans using the breeding populations normally used within the NDSU dry bean breeding program.

Specific objectives are:

- To identify genomic regions involved in controlling four agronomic traits of economic importance i.e. seed yield, maturity, 100- seed weight and plant weight.
- To identify possible candidate genes with known functions which fall within the significant QTL regions.

MATERIAL AND METHODS

This part presents materials and methods followed for this GWAS and consist of three different sections. Section one provides detailed information on the phenotypic analysis. Section two describes genotypic analysis, and section three explains the procedure followed for association analysis.

Phenotypic analysis

Experimental field (phenotypic) data

Phenotypic data was collected from breeding lines of the NDSU dry bean breeding program from the AYTs (Advanced Yield Trials) and PYTs (Preliminary Yield Trials). These trials were grown in year 2012 at four locations (Carrington, Hatton, Prosper and Johnstown in North Dakota). Genotypes tested at four contrasting locations were selected to have a more accurate estimate of GXE interactions. The trials represent the Mesoamerican gene pool, in which race Durango includes the market classes' pinto, pink, red, and great northern while race Mesoamerican includes market classes blacks and navy. Depending upon the number of genotypes to test, the experimental designs used at all sites were either partially balanced square lattices with two replicates or a randomized complete block design with three replicates.

Phenotypic data for many traits are routinely collected in a breeding program, but for the purposes of this study only four important agronomic traits were selected: (1.) Days of maturity (d) which is the actual number of days after sowing when approximately 95% of plants in a plot have at least one dry pod. (2.) 100-seed weight (g), which is measured by weighting 100 randomly undamaged seeds and recorded in grams at approximately 10% moisture. (3.) Plant height (cm) is measured from soil surface to the top node bearing at least one dry pod with seed.

(4.) Seed yield (kg ha^{-1}) at standardized to16% moisture and rounded up to nearest whole number.

These data represent the agronomic performance in a wide range of environments (Table 1). All lines from AYT+PYT (208) were divided into different subgroups of AYT, PYT, Durango and Mesoamerican and were considered as different populations for the GWAS. ANOVA was be performed using SAS Proc GLM/MIXED (SAS Institute, Inc. 2011) (Freund *et al*., 1986), to see if genotypes were significantly different and also to calculate least square means (to reduce errors in the replications) and standard deviations for phenotypic data evaluation from each trial, after discarding the missing data. Replications and environments were considered as random effects while genotypes are fixed. Since genotypes grown were not always common over all locations, no further statistical analysis was performed.

Genotypic analysis

Plant material

For genomic DNA extraction, all genotypes were planted in the greenhouse. After two to three weeks of growth, the first trifoliate leaves were harvested and stored at –80º C prior to extraction. DNA was extracted using cetyltrimethyl ammonium bromide (CTAB) method (Doyle and Doyle, 1987; Doyle and Doyle, 1990). A total of 50 µl of extracted DNA, with concentration of 100 ng μl^{-1} was sent to USDA-ARS Soybean Genomics and Improvement Lab Facilities at Beltsville-MD for the amplification of SNPs (Hyten *et al*., 2010). All the samples were screened to generate genotypic data that was generated using 6K SNP chip (designed using the Illumina Infinium system). BeanCAP 6k chip has been used in this study. Two Illumina iSelect Genechips for SNP analysis was developed by Perry Cregan at USDA-ARS, Beltville, MD. For this development, 19.6 billion bases of sequence data from 19 diverse common bean genotypes was

obtained, from which 10,453 SNPs were obtained. It was estimated that 6533 SNPs are segregating in 288 F_2 line of the NDSU Stampede x Red Hawk mapping population. These SNP markers are evenly distributed across the 11 chromosomes for both advanced lines as well as from the preliminary lines (BeanCAP, 2014). The genomic data was obtained using genome studio (Illumina, 2014).

Table 1. Total number of genotypes by market class and by sub-population.

Population	Black	Great Northern	Navy	Pinto	Red and Pinks	Total
AYT	46	17	36	21	16	136
PYT	_	_	_	38	34	72
Durango	_	17	_	59	50	126
Mesoamerican	46	_	36	_	_	82
AYT+PYT	46	17	36	59	50	208

All lines from AYT+PYT (208) were divided into different subgroups of AYT, PYT, Durango and Mesoamerican and were considered as different populations for the GWAS.

Association studies

Imputation of missing genotypes and minor allele frequency

For imputation, FastPHASE v. 1.3 (Scheet and Stephens, 2006) software was used. This software imputed missing genotypes for unmeasured SNPs using an estimated maximum likelihood algorithm using default parameters. Minor allele frequency (MAF) was calculated separately for advanced yield trails, preliminary yield trials, Durango, Mesoamerican and combining all genotypes together. Markers with MAF < 0.05 were removed and remaining polymorphic markers were used for further analysis in each of the population analyzed.

Principal component analysis

Different approaches were used to assess population structure, family relatedness, and structured association. The purpose of using these approaches was to remove all the false positives which were generated in the population structure by using genotypic information of all

the markers in the genome. To control population structure and relatedness among all individuals of different subsets of breeding population, Principal Component Analysis (PCA) was done using the PRINCOMP procedure of SAS 9.3 (SAS Institute, Inc., 2011). This procedure is done to convert correlated variables into smaller number of uncorrelated variables, where each successive component explains for decreasing amount of data variation (Carsten *et al*., 2012). Principal components that explained cumulative of at least 25% and 50% of variation were selected for successive analyses to control for population structure.

Linear regression models

To identify marker-trait associations, six different models were implemented using the MIXED Procedure in SAS 9.1.3 (SAS Institute, Inc., 2011) and Gemma (Zhou and Stephens, 2012). Of all the regression models, Naïve, Principal component (PC 25%), Principal component (PC 50 %) estimated in SAS 9.3 where as other models, Kinship (K), Principal component (PC 25%) + Kinship, Principal component (PC 50%) + Kinship were estimated in Gemma (Mamidi *et al.*, 2011) . All general linear models (GLM) considered only fixed effects and all mixed linear models (MLM) considered both fixed and random effects (Table 2).

Table 2. Statistical models used for marker-trait association.

Model		Approach
Naïve	$y = X\alpha + \varepsilon$	GLM
PC (25% Variance explained)	$y = X\alpha + P\beta + \varepsilon$	GLM
PC (50% Variance explained)	$y = X\alpha + P\beta + \varepsilon$	GLM
K	$y = X\alpha + K\nu + \varepsilon$	MLM
K+PC (25% Variance explained)	$y = X\alpha + P\beta + K\nu + \varepsilon$	MLM
K+PC (50% Variance explained)	$y = X\alpha + P\beta + K\nu + \varepsilon$	MLM

Where, y is a vector for the phenotypic observation, α is a vector for the fixed effects of the SNPs, β is a vector for the fixed effects of the population structure, ν is a vector for the random effects of the individual relatedness, and ε is a vector for the residual effects (Mamidi *et al*., 2011). X is the SNP genotypes, P is the matrix of the principal components (PCs), K is K-matrix, GLM is general linear models and MLM is mixed linear models.

For each model, expected P- values were calculated by dividing rank of observed P-values which was ranked in order from smallest to largest. Mean square difference (MSD) was calculated to find the deviation of observed P-value from this uniform distribution (Mamidi *et al*. 2011). Best model out of six models that controls for population structure and reduces false positive identification was selected based on the MSD values. The model which showed less deviation of P value or has least MSD was selected as the best model for genome wide association analysis (Mamidi *et al*., 2011).

Significant SNP markers ($p< 0.001$) were selected from the selected best models and Manhattan plots were constructed using –log10 of p-values against chromosome location with SAS 9.3 (SAS Institute, Inc. 2011), to represent position of these markers on chromosome graphically. The phenotypic variation (R^2) and the allelic means were calculated for each of the significant markers using the Reg and Means procedures in SAS 9.3 (SAS Institute, Inc., 2011). To select markers which show higher level of significance even in the presence of other significant markers, the stepwise regression analysis was performed, using SAS with significance level for entry as 0.05 and significance level for staying as 0.05. Positive false discovery rate (pFDR; Q values) was calculated for each trait using the PROC MULTTEST procedure in SAS 9.3 (SAS Institute, Inc. 2011).

Many different allelic combinations were detected associated with each trait using data from genotypes and phenotypes of all populations. The cut off values were based on the mean $\pm$ standard deviation of controls (Table 3). Alleles of all significant markers were combined to make combinations. These combinations could be used for selection of genotypes with trait of interest in further studies.

Table 3. Cut off values for each trait in all populations.

Populations		100-Seed weight	Maturity	Yield	Plant height
AYT, PYT, AYT+PYT	Mean	31	102	2600	55
	Standard deviation	10	11	441	4
	Mean + SD	40	112	3041	59
	Mean - SD	21	91	2159	51
Mesoamerica	Mean	21	102	2600	54
	Standard deviation	1	11	441	3
	Mean + SD	22	112	3041	57
	Mean - SD	20	91	2159	51
Durango	Mean	40	102	2600	56
	Standard deviation	3	11	441	5
	Mean + SD	42	112	3041	61
	Mean - SD	37	91	2159	51

QTL regions and candidate genes

Significant markers that were present in more than one population were identified for each trait. Markers in linkage disequilibrium with these significant markers that were identified using TASSEL to define QTL regions. Heat maps were constructed with markers that fall within the QTL region using TASSEL, and to find the markers that are in LD (R^2 value > 0.5).

Genes close to significant SNP markers found to be associated with plant height, seed weight, seed yield and maturity were obtained from the *Phaseolus vulgaris* annotation (Schmutz *et al*. 2014) available at phytozome.net. Genes that were the present within the physical position (Mb) of QTL region were selected. Physiology and functions of these genes were studied and only those genes that were previously reported to be responsible for any these traits in any other crop or plant were considered as candidate genes. Presence of the candidate genes in previous studies confirmed the consistency that they are significantly associated with the traits of interest.

RESULTS

Phenotypic analysis

Phenotypic data for each trait was evaluated in the bean breeding program from different locations. After taking the least square means of each genotype from each trial and combining across locations, days of maturity ranged from 75 to 113 days and 100-seed weight , was in a ranged between 18 to 45 g. The range for plant height observed across all the genotypes was from 41 cm to 69 cm. Seed yield was obtained between 1394 kg ha^{-1} to 3565 kg ha^{-1} (Table 4).

Table 4. Means and range for all traits after discarding the missing data.

Trait	Mean	Standard deviation	Min	Max
100-seed weight	30	8	18	45
Maturity	96	13	75	113
Yield	2543	409	1394	3565
Plant Height	56	4	41	69

DNA was extracted from all the lines. Genotypes with missing genotypic data or missing phenotypic data were discarded and a total of 208 genotypes were used for GWAS. As mentioned before, these lines were subdivided into five subpopulations based either on their breeding stage or the race they belong to: AYT, PYT, Mesoamerican, Durango and AYT+PYT.

Polymorphic marker analysis

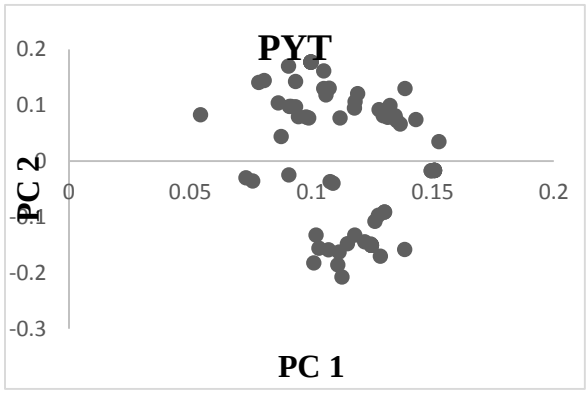
Polymorphic markers were selected for all the genotypes based of the distribution in minor allele frequencies (MAF). Across all subpopulations, SNP on 6K chip showed MAF distribution ranging from 0 to 50%. Out of all SNP, only informative markers with MAF > 5 % were included for succeeding analyses (Table 5).

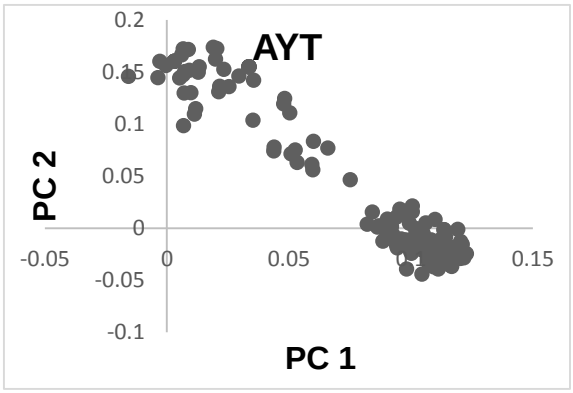
Table 5. Number of polymorphic markers (MAF > 0.05) used for analyses.

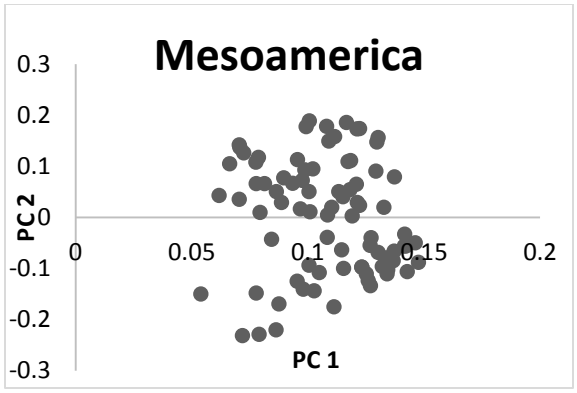
Genotypes	Number of genotypes	Number of Markers
Advanced Yield Trial (AYT)	136	2958
Preliminary Yield Trial (PYT)	72	2816
Durango	126	2851
Mesoamerican	82	1943
AYT + PYT	208	3046

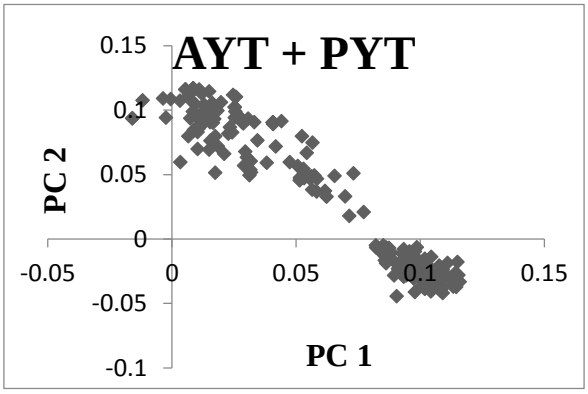
Population structure and kinship analysis

To estimate the population structure, PCA was implemented. For subpopulation AYT + PYT, PYT, and Durango, first 3 principal components PC explained 50% of total variance while for Mesoamerican and AYT, first 2 principal components (PC) explained 50% of total variance. The graphs that were plotted using first two PCs, explained the distribution of genotypes within different subpopulations (Figure 1).









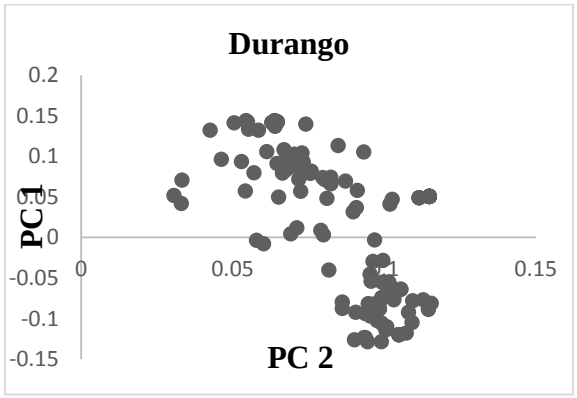


Figure 1. Principal component analysis of all individual genotypes present in five sub-populations of beans. The x-axis represents the eigenvalue for principal component 1 (PC1) and the y-axis represents the eigenvalue for principal component 2 (PC2).

Marker-trait associations

To analyze the association between single nucleotide polymorphic markers with the agronomic traits, and to reduce confounding effect of population structure, six statistical models were selected (Table 2). The marker-trait analysis was conducted using both phenotypic data collected from field (2012) and genotypic data. After comparing the MSD values of each model, best model with least MSD was selected for each trait (Mamidi *et al*., 2011) (Table 6).

Table 6. Test statistics for the six models with MSD values used for association analysis in five different subpopulations for plant height, seed weight, seed yield and maturity.

	Traits	Naïve	PC25	PC50	Kinship	PC25+ Kinship	PC50+ Kinship
PYT+AYT	Plant Height	0.0983	0.0451	0.0352	**0.0001**†	0.0002	0.0004
	100-Seed Weight	0.2489	0.0436	0.0368	0.0018	**0.0002**†	0.0006
	Maturity	0.2161	0.0308	0.0397	0.0029	0.0023	**0.0022**†
	Seed Yield	0.0117	0.0153	0.0282	0.0002	0.0003	**0.0001**†
PYT	Plant Height	0.0122	0.0120	0.0114	**0.0003**†	0.0004	0.0004
	100-Seed Weight	0.1085	0.0971	0.0971	0.00032	**0.00031**†	0.0017
	Seed Yield	0.0726	0.0394	0.0158	0.0011	0.0008	**0.0007**†
AYT	Plant Height	0.0174	0.0466	0.0314	**0.0005**†	0.0007	0.0024
	100-Seed Weight	0.2374	0.0111	0.0110	0.0024	**0.0001**†	0.0004
	Maturity	0.2141	0.0302	0.0299	0.0026	**0.0019**†	0.0021
	Seed Yield	0.1247	0.0370	0.0189	**0.0002**†	0.0002	0.0002
Durango	Plant Height	0.0531	0.0567	0.0288	0.0004	0.0005	**0.0004**†
	100-Seed Weight	0.1129	0.0563	0.0327	0.0013	0.0011	**0.0003**†
	Maturity	0.0739	0.0311	0.0063	0.0005	**0.0003**†	0.0005
	Seed Yield	0.0081	0.0012	0.0012	0.0013	0.0012	**0.0012**†
Mesoamerica	Plant Height	0.0467	0.0420	0.0388	0.0414	**0.0367**†	0.0380
	100-Seed Weight	**0.0182**†	0.0189	0.0250	0.0199	0.0344	0.0392
	Maturity	0.0336	0.0348	**0.0183**†	0.0491	0.0503	0.0465
	Seed Yield	0.0241	0.0248	**0.0164**†	0.0342	0.0329	0.0320

† Least mean square deviation values (MSD) among all models selected for GWAS. PC 25% and PC 50% - principal component analysis with at least 25% and 50% variance respectively. PC+K=PCA + Kinship.

For plant height, Kinship model was selected for subpopulations AYT + PYT, PYT and AYT and models K + PC 25% and K + PC 50% were selected for subpopulations Mesoamerica

and Durango. For 100-seed weight, PC 25% +K (PYT + AYT, AYT) and K+ PC 50% (Durango) and Naïve (Mesoamerican) showed the best MSD values. Regression Model PC 50% + K (PYT+AYT, PYT and Durango), K (AYT) and K (Mesoamerican) were selected for seed yield. For the trait maturity, PC 25% for Mesoamerica, PC 25% +K for AYT and Durango and PC 50% + K for AYT + PYT were selected as the best model.

Thus, subpopulations AYT+ PYT, PYT, AYT, and Durango, have kinship (K) or both kinship and population structure (K+PC 25%, K+PC 50%), as the best model while for Mesoamerican subpopulation, except for plant height, the naïve, PC 25% and PC 50% were the best model selected. The Q-Q plots showed the distribution of observed P- values and expected p-values for all the subpopulations (Figure 2-6).

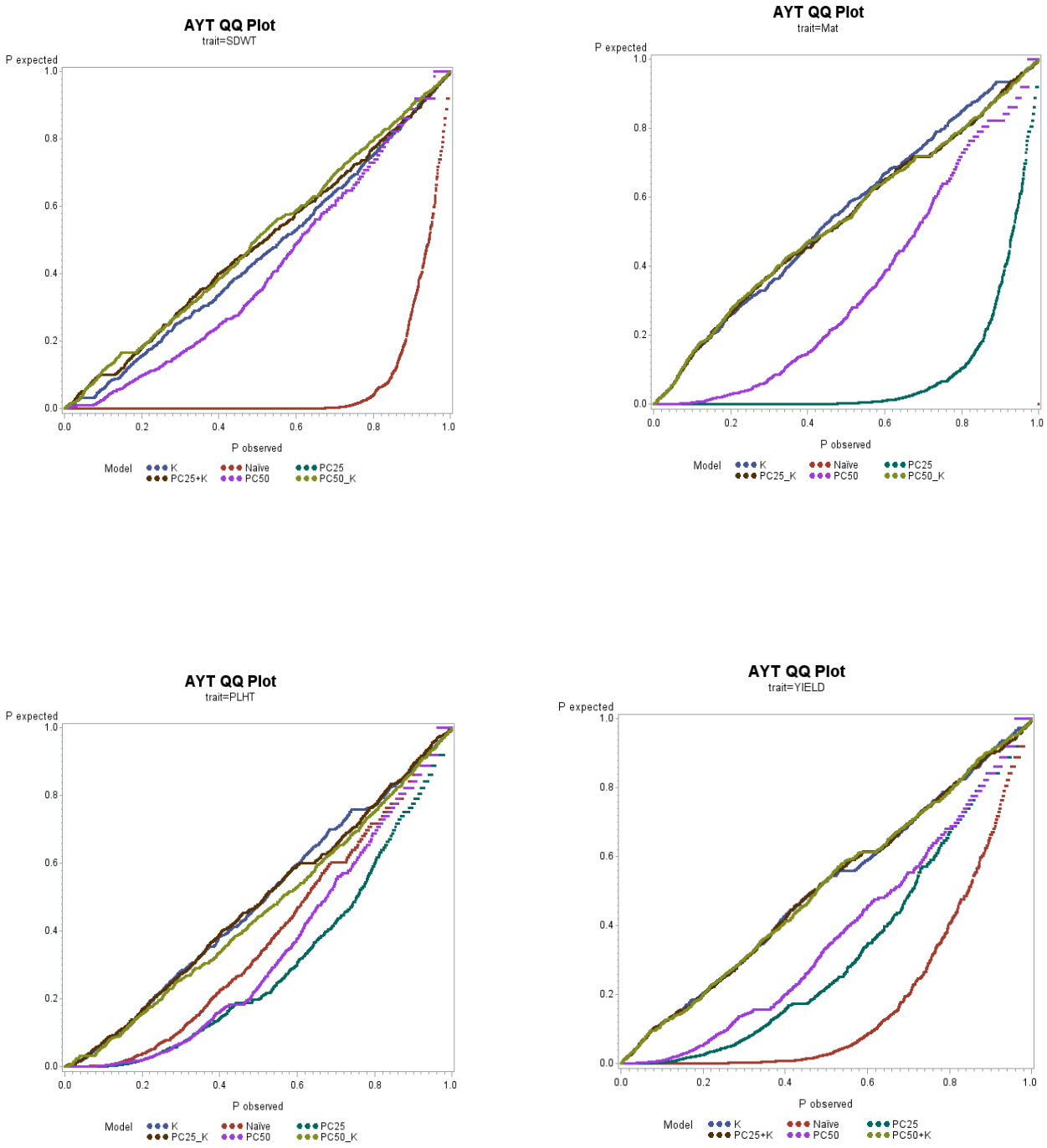


Figure 2. Graphical representation of distribution of p-values for six models for subpopulation AYT.

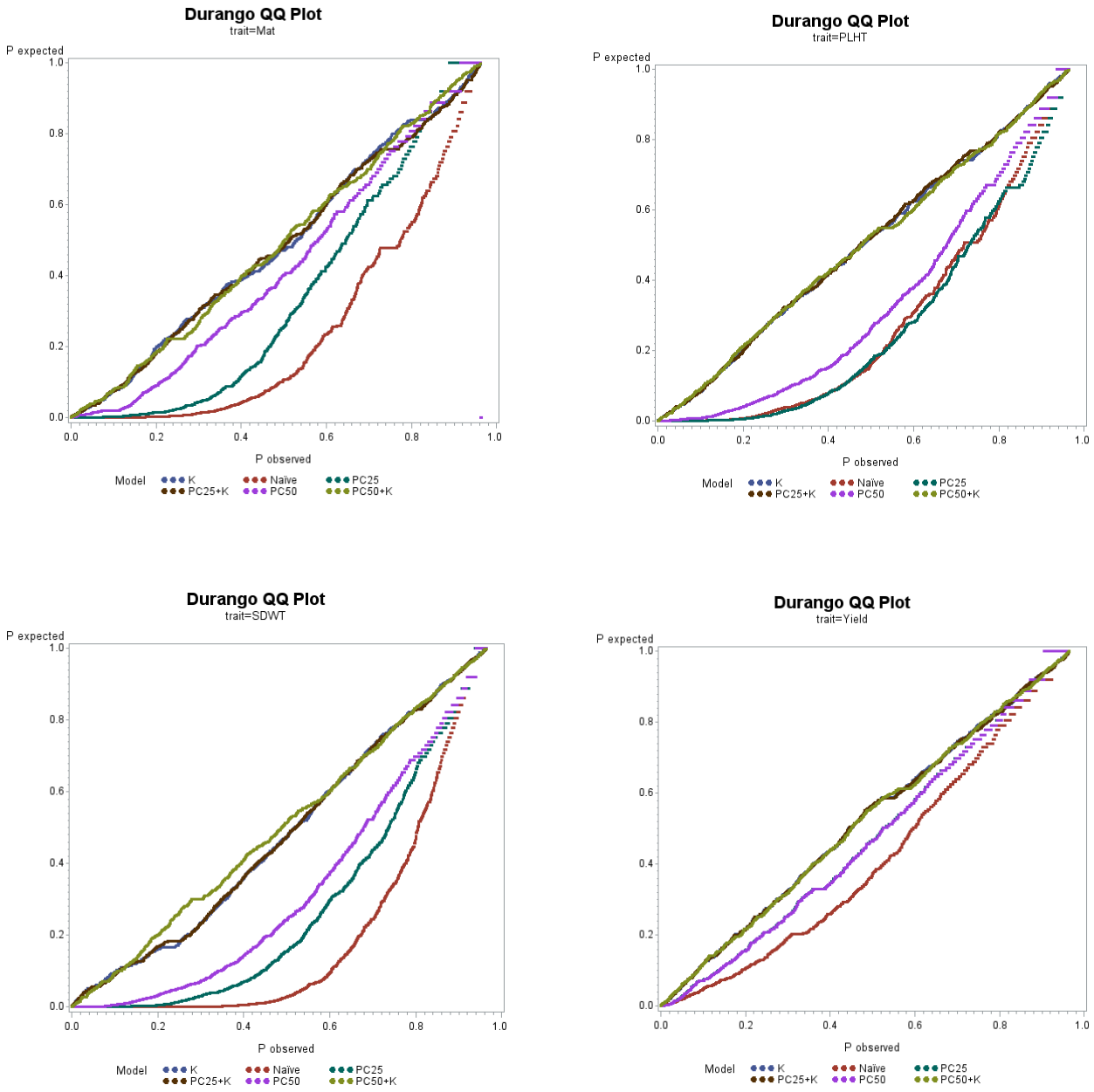


Figure 3. Graphical representation of distribution of p-values for six models for subpopulation Durango.

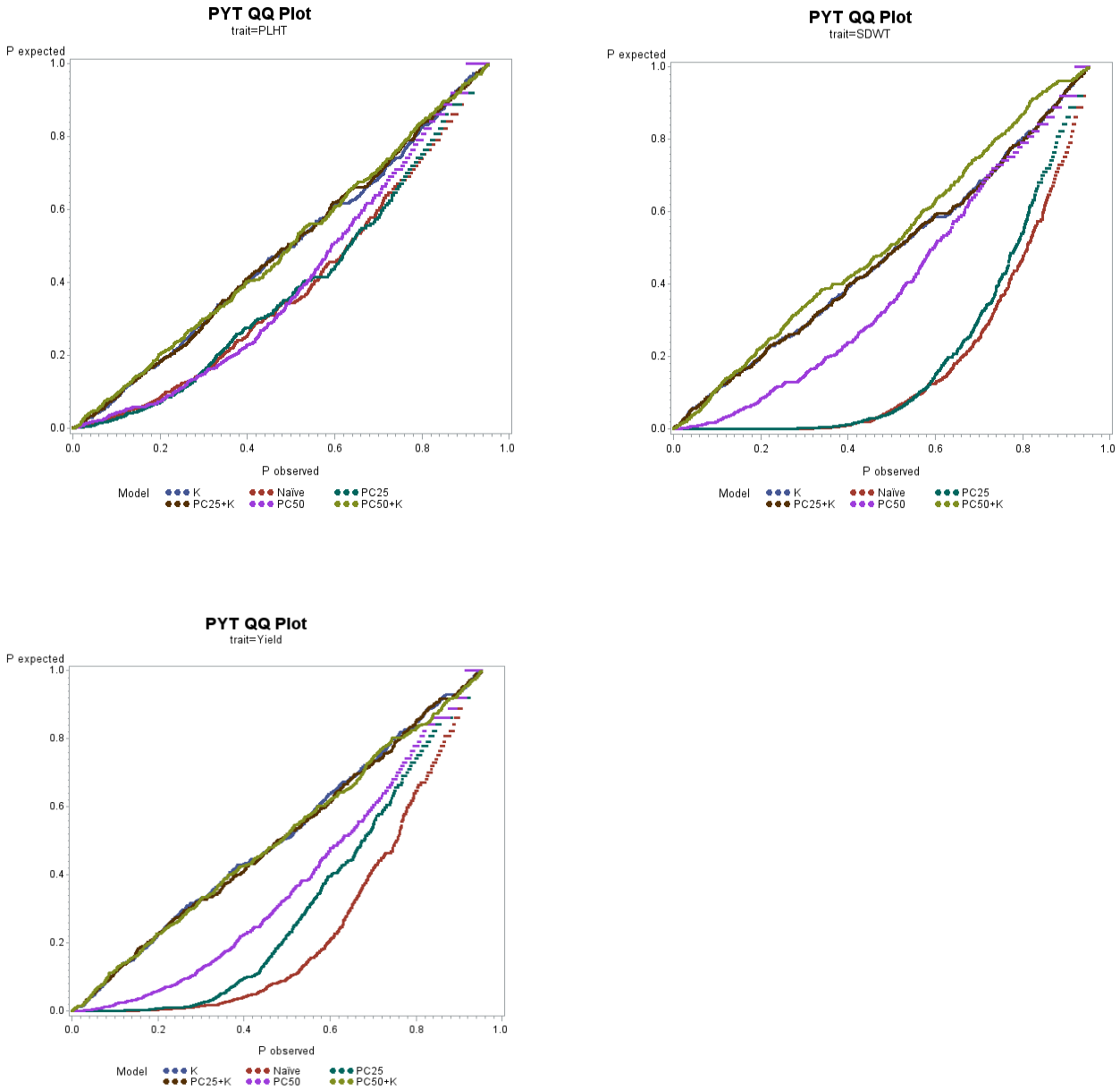


Figure 4. Graphical representation of distribution of p-values for six models for subpopulation PYT.

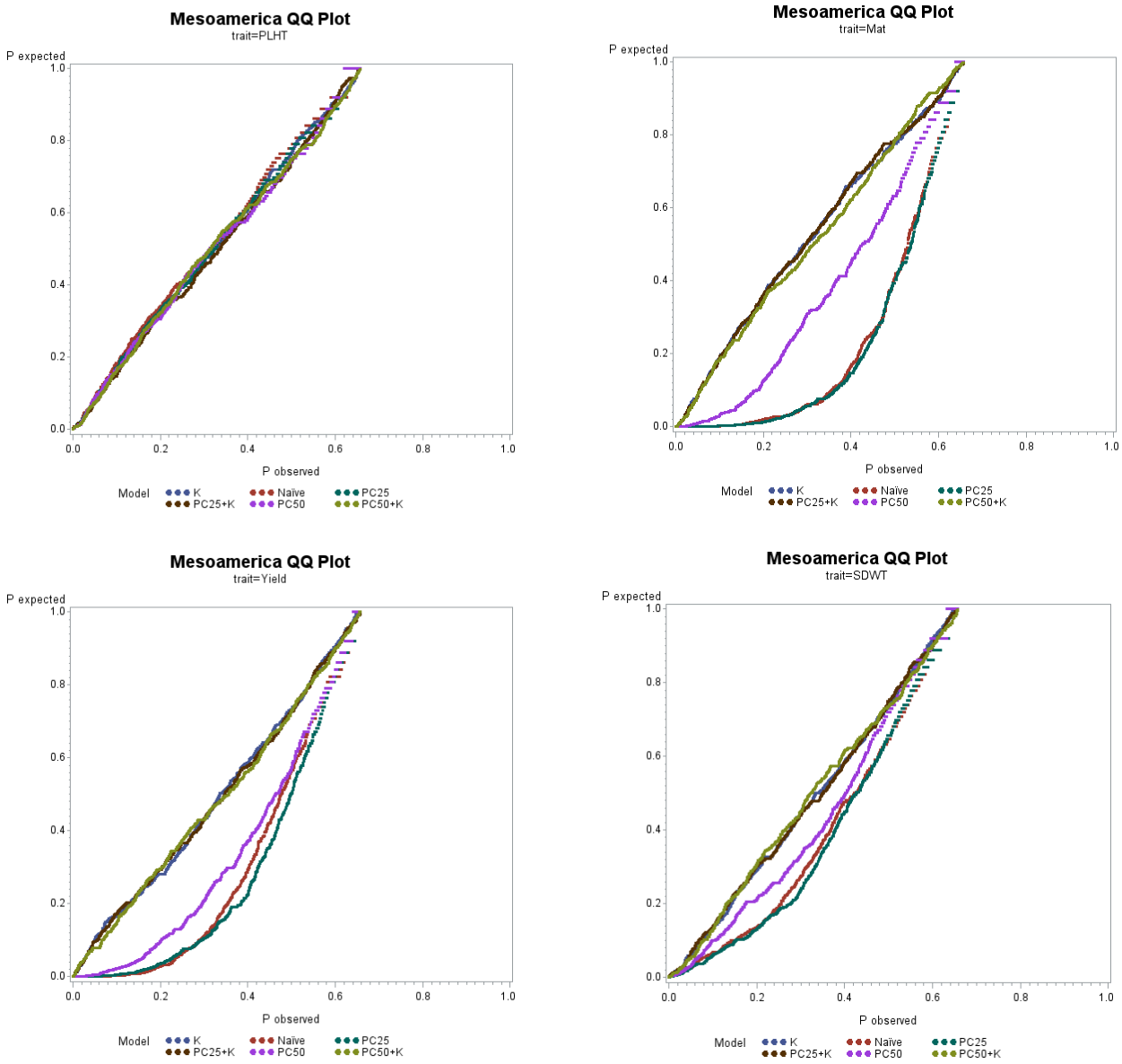


Figure 5. Graphical representation of distribution of p-values for six models for subpopulation Mesoamerican.

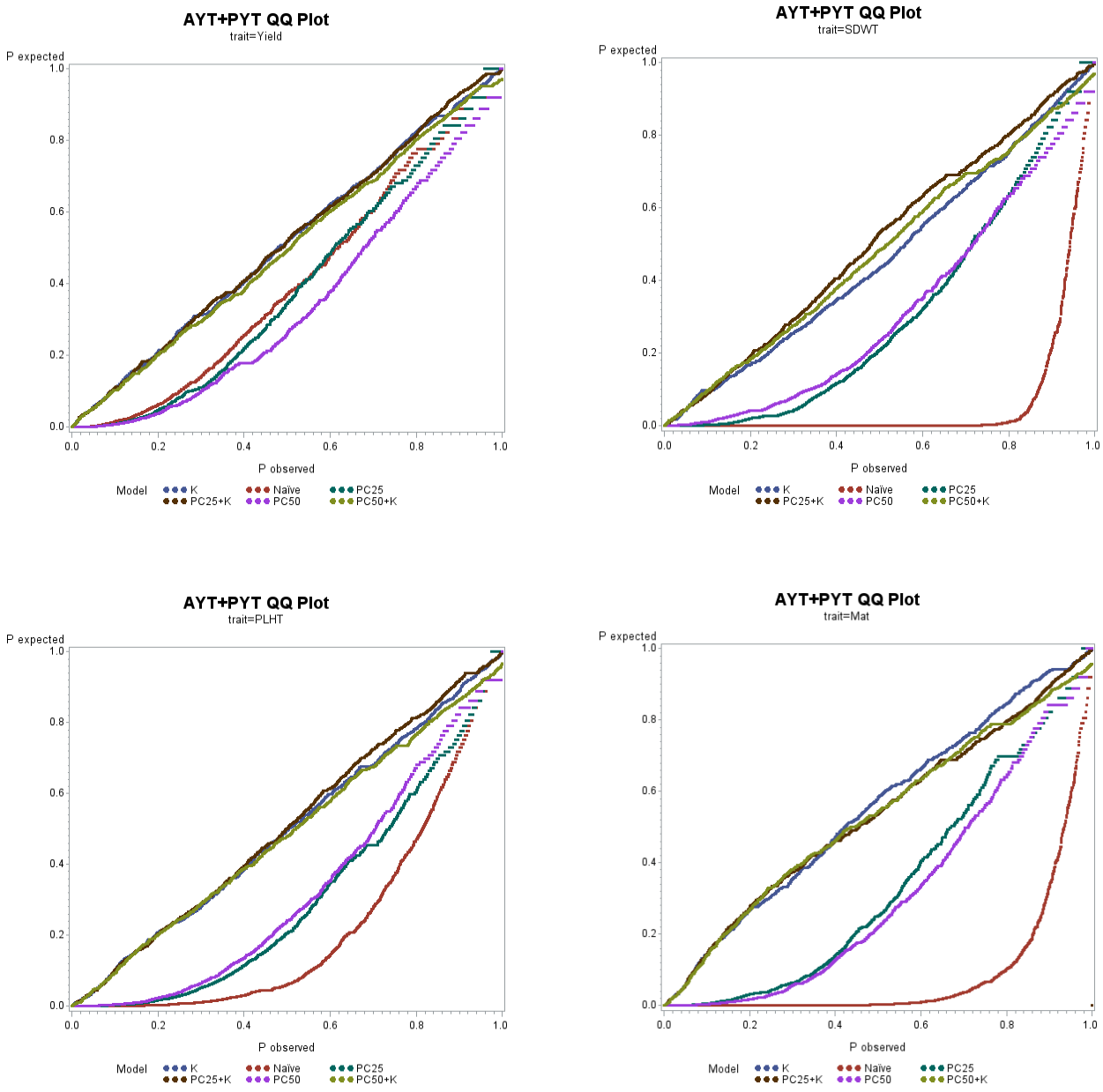


Figure 6. Graphical representation of distribution of p-values for six models for subpopulation AYT +PYT.

Significant markers

Significant markers were identified from the best models selected for each trait for all subpopulations by performing simple regression model. Markers which met the cut off of $p < 0.001$ or ($-\log10 (P) > 3.0$) and pFDR (Q value) < 0.1, were selected as significantly associated with traits.

Manhattan plots were made to represent the chromosomal position of significant markers for each trait. The plots were made using -log10 of p-values against chromosome location. Different colors present different chromosomes and black line is to show the Bonferroni corrected significance threshold (Figure 7-10).

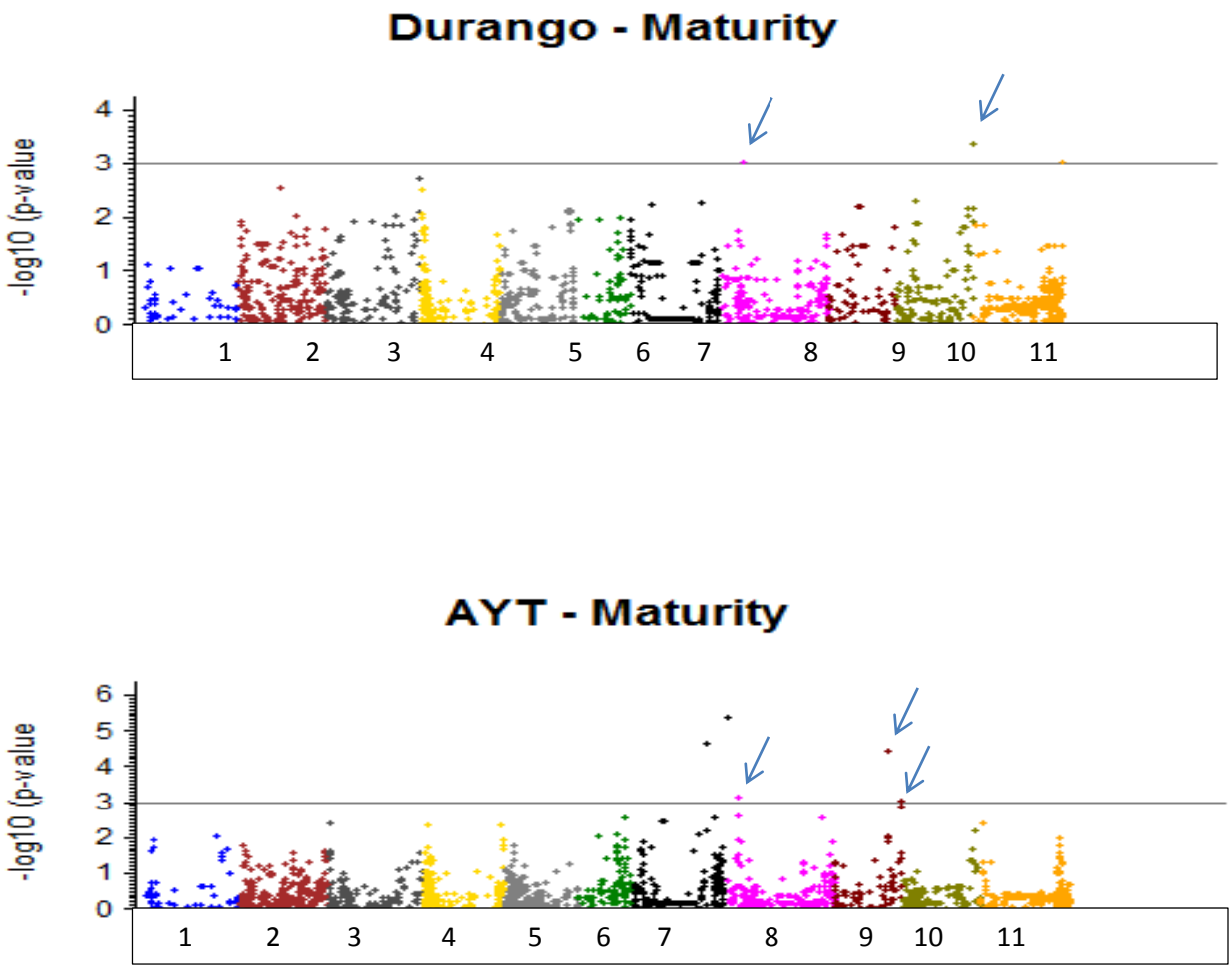


Figure 7. Manhattan plots of representing markers associated with maturity. The black line shows the p-value that corresponds FDR of 0.05. Blue arrows point out most significant markers after stepwise REG. Each color represents different chromosomes ranging from 1 to 11.

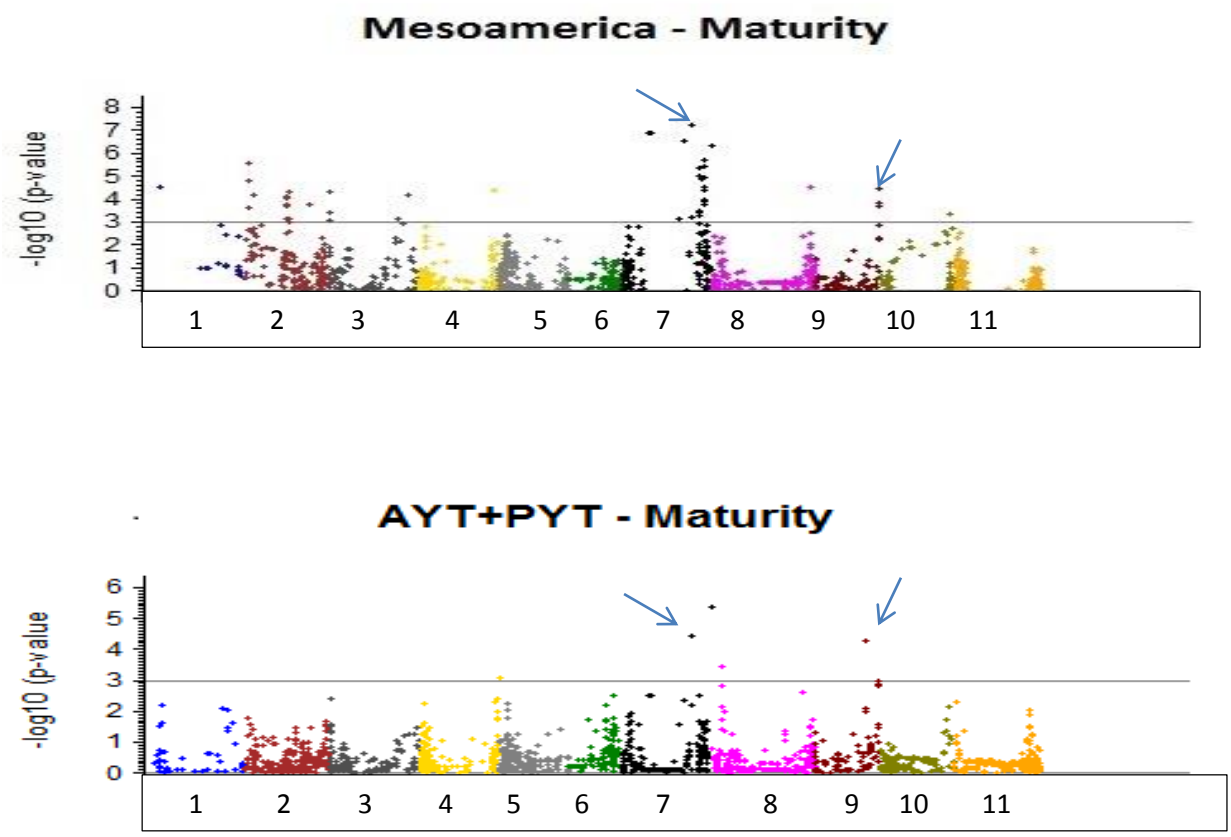


Figure 7. Manhattan plots of representing markers associated with maturity. The black line shows the p-value that corresponds FDR of 0.05. Blue arrows point out most significant markers after stepwise REG. Each color represents different chromosomes ranging from 1 to 11 (continued).

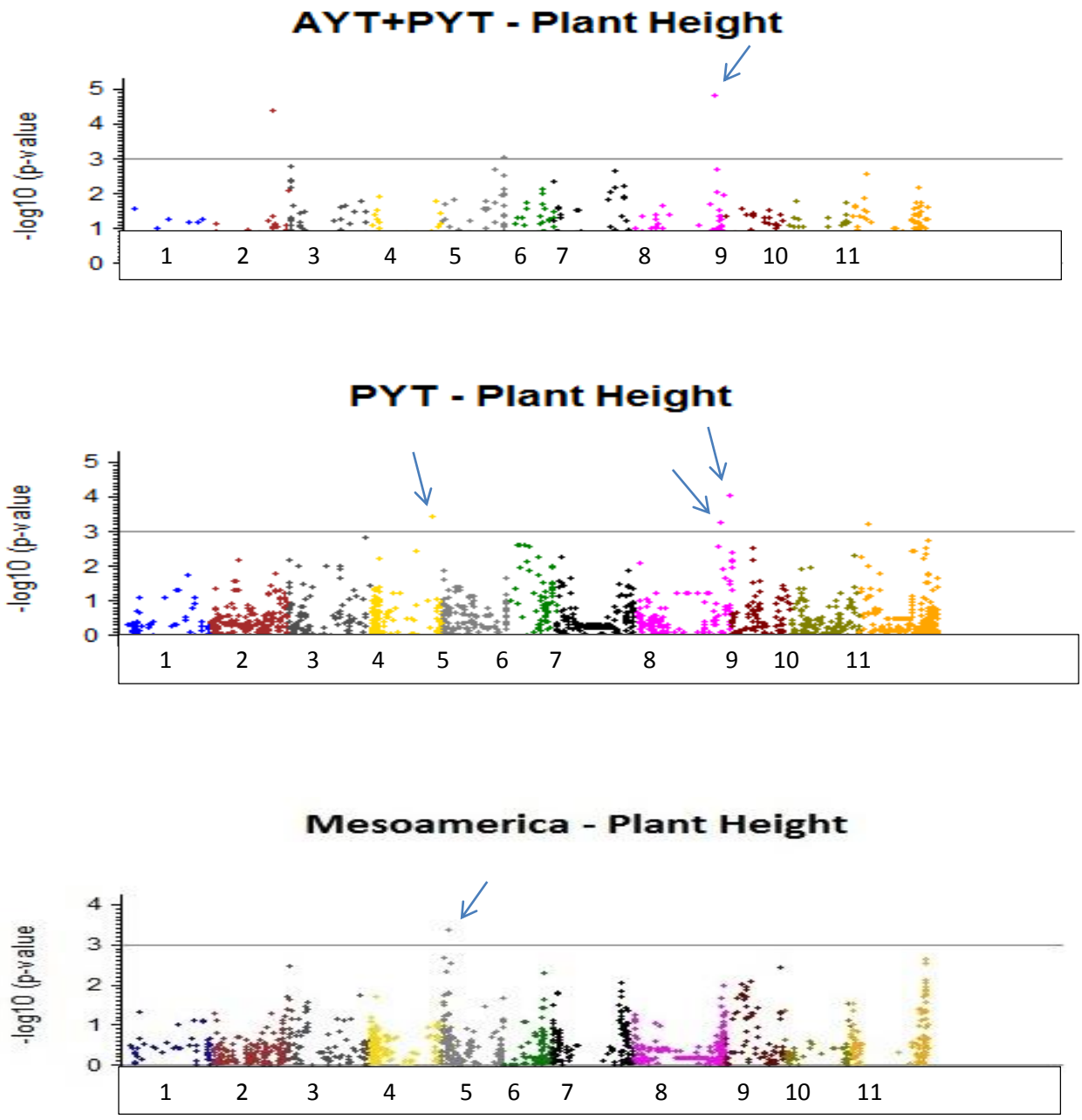


Figure 8. Manhattan plots of representing markers associated with plant height. The black line shows the p-value that corresponds FDR of 0.05. Blue arrows point out most significant markers after stepwise REG. Each color represents different chromosomes ranging from 1 to 11.

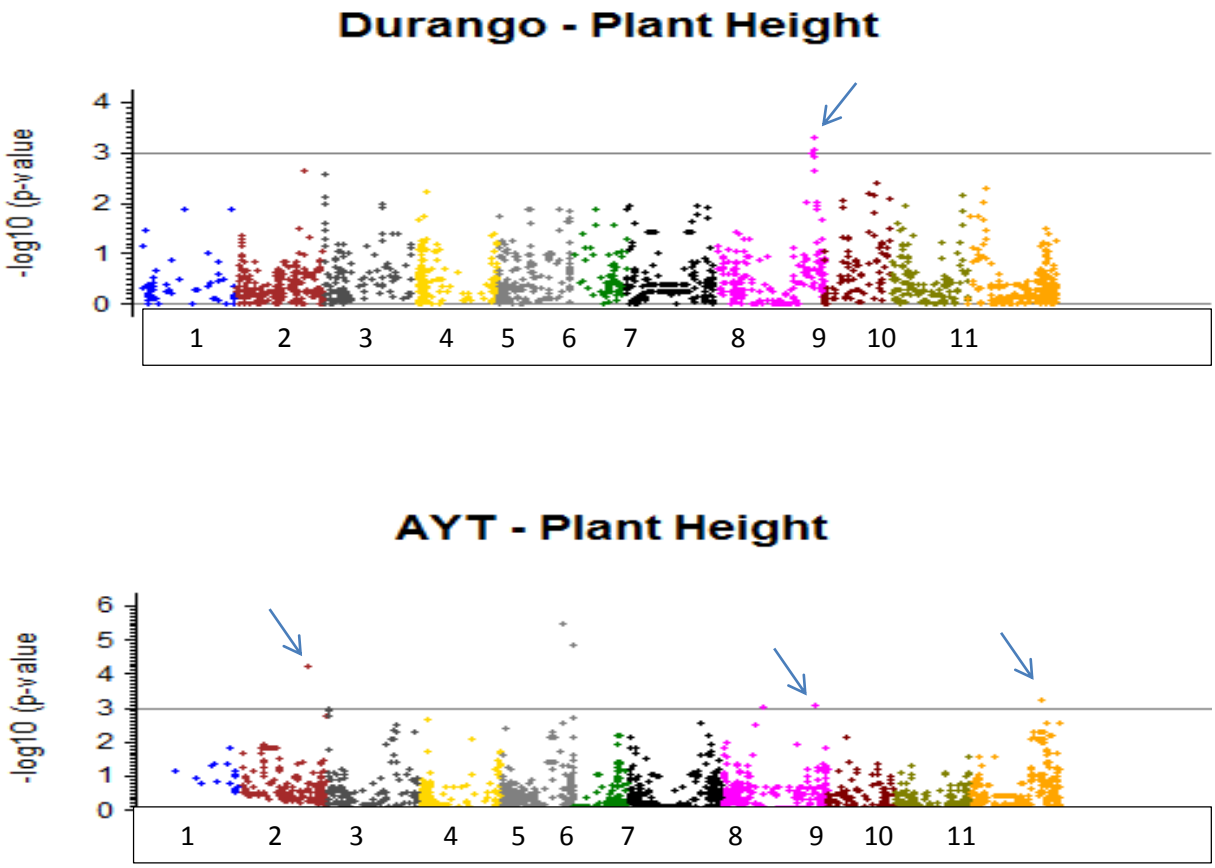


Figure 8. Manhattan plots of representing markers associated with plant height. The black line shows the p-value that corresponds FDR of 0.05. Blue arrows point out most significant markers after stepwise REG. Each color represents different chromosomes ranging from 1 to 11 (continued).

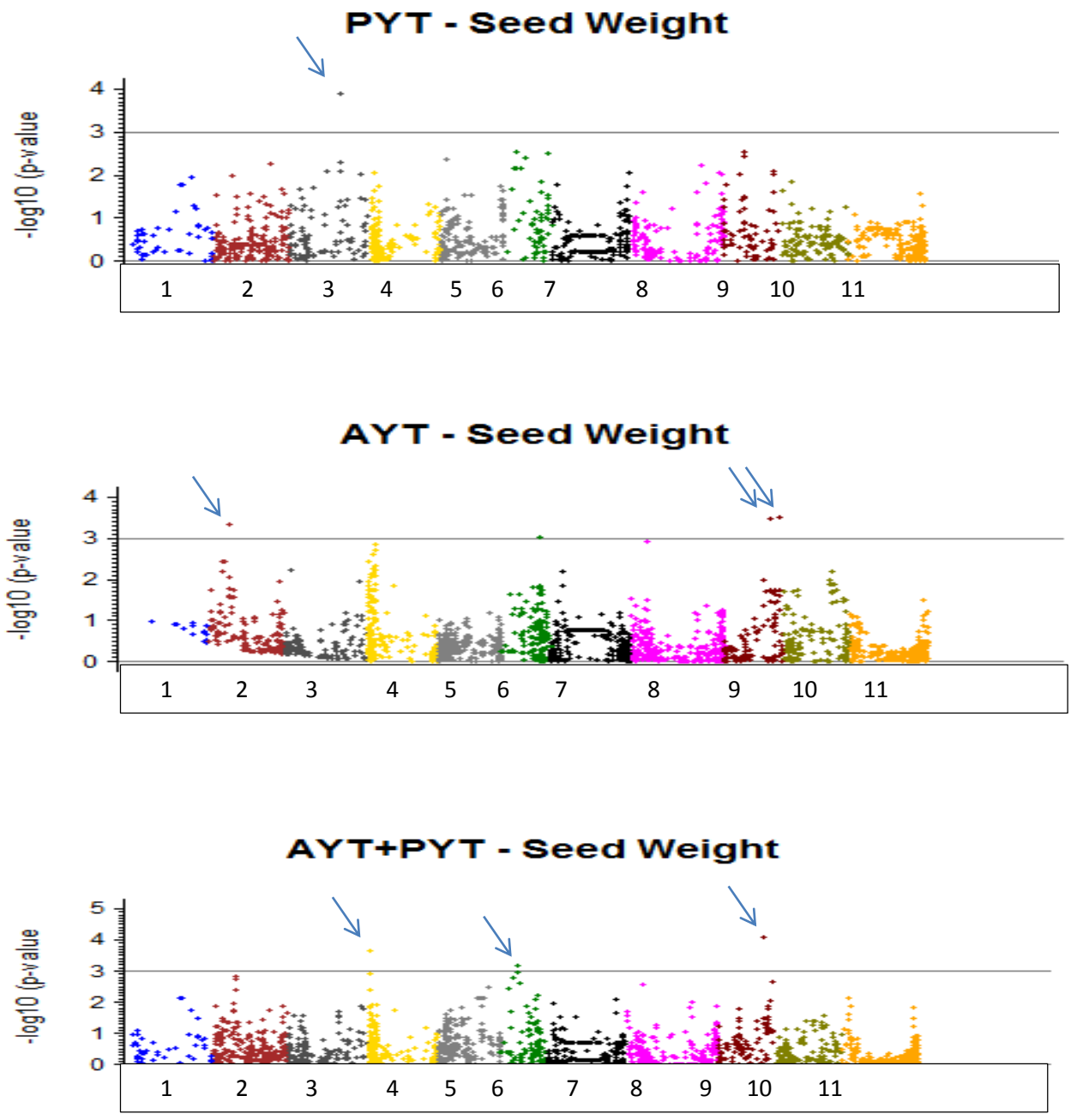


Figure 9. Manhattan plots of representing markers associated with 100-seed weight. The black line shows the p-value that corresponds FDR of 0.05. Blue arrows point out most significant markers after stepwise REG. Each color represents different chromosomes ranging from 1 to 11.

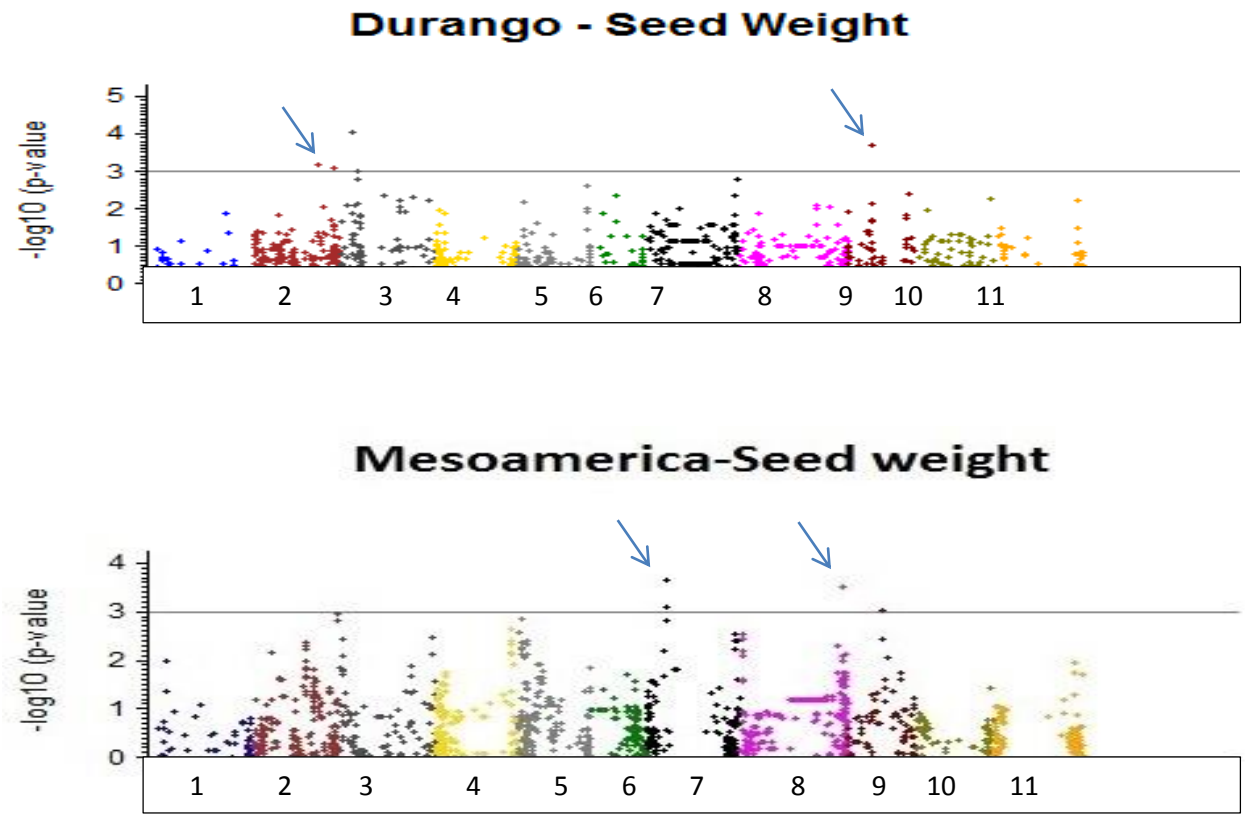


Figure 9. Manhattan plots of representing markers associated with 100-seed weight. The black line shows the p-value that corresponds FDR of 0.05. Blue arrows point out most significant markers after stepwise REG. Each color represents different chromosomes ranging from 1 to 11 (continued).

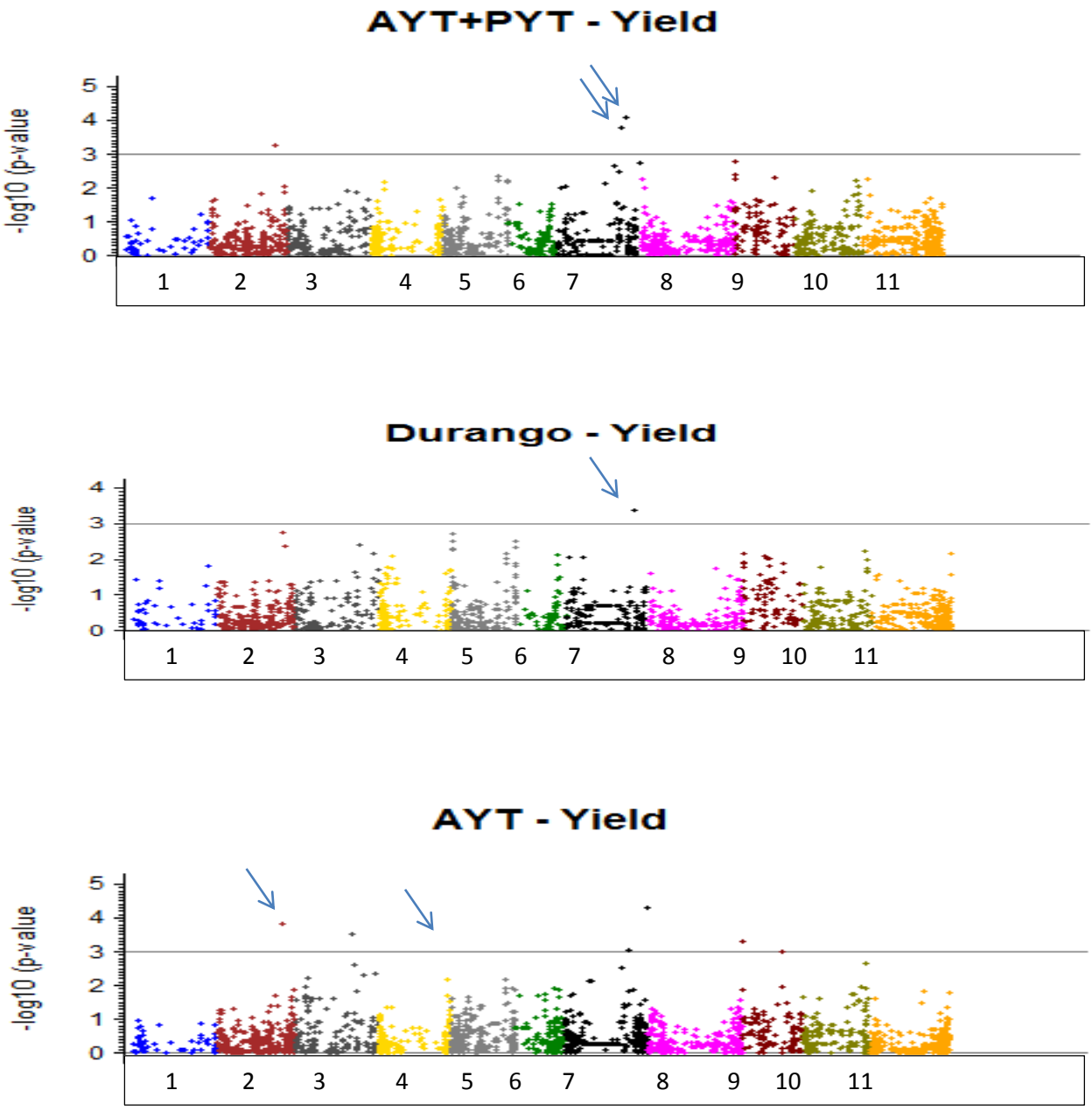


Figure 10. Manhattan plots of representing markers associated with seed yield. The black line shows the p-value that corresponds FDR of 0.05. Blue arrows point out most significant markers after stepwise REG. Each color represents different chromosomes ranging from 1 to 11.

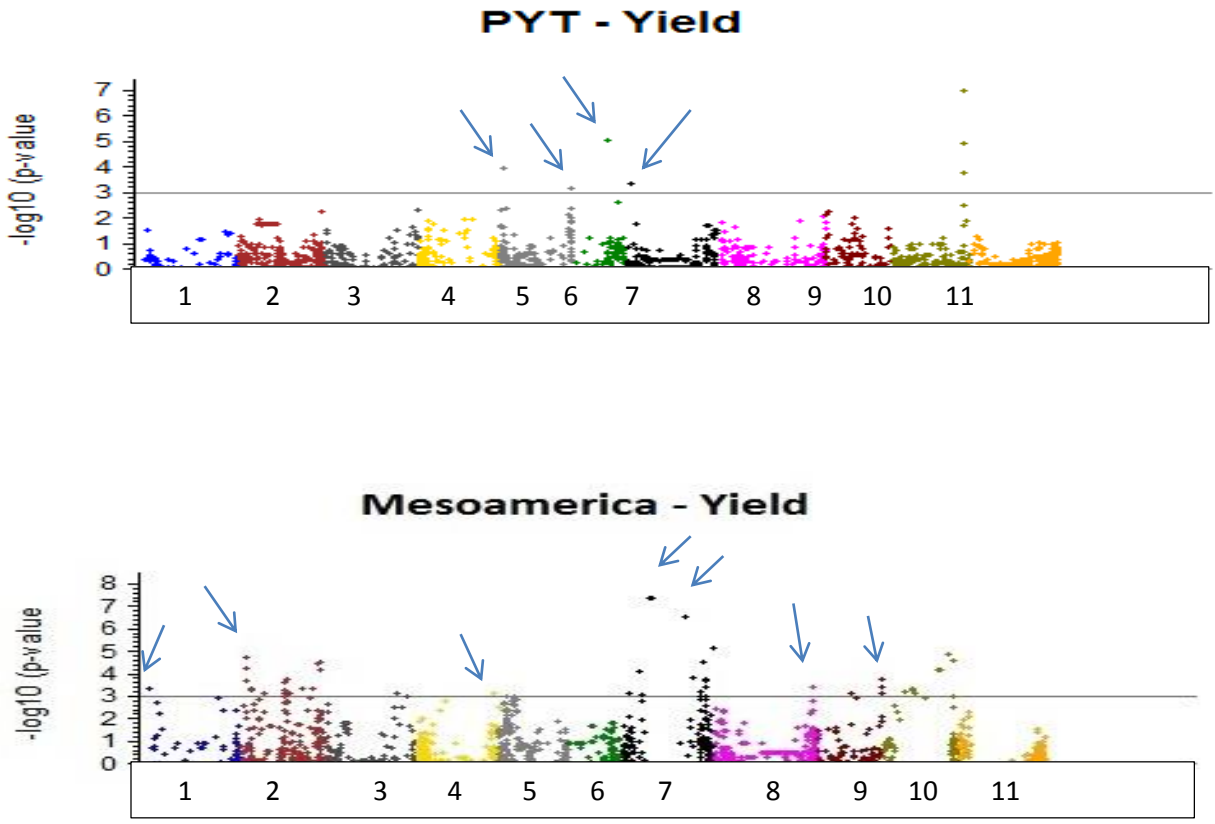


Figure 10. Manhattan plots of representing markers associated with seed yield. The black line shows the p-value that corresponds FDR of 0.05. Blue arrows point out most significant markers after stepwise REG. Each color represents different chromosomes ranging from 1 to 11 (continued).

In Reg and Means procedure, a total of 6, 7, 4, 1, and 73 markers were found to be associated with seed yield, 8, 4, 4, 4, and 1 markers for plant height and 4, 1, 4, 7, and 5 markers with 100-seed weight for subpopulations AYT, PYT, AYT+PYT, Durango, and Mesoamerican respectively. For maturity, 7, 7, 3 and 66 markers for AYT, AYT+PYT, Durango, and Mesoamerica, respectively were found to be significantly associated. After stepwise regression analysis, a reduction in significant markers was observed for all traits (Table 7).

Table 7. Number of significant markers with R-Square values included in stepwise regression model.

	Traits	No. of initial markers	No. of markers included after stepwise regression	R^2
AYT	Seed Yield	6	2	12.32%
	Plant Height	8	2	13.62%
	Maturity	7	3	27.25%
	100-Seed Weight	4	3	65.14%

	Traits	No.of initial markers	No. of markers included after stepwise regression	R^2
PYT	Seed Yield	7	4	62.38%
	Plant Height	4	3	36.76%
	100-Seed Weight	1	1	01.23%

	Traits	No.of initial markers	No. of markers included after stepwise regression	R^2
AYT + PYT	Seed Yield	4	2	19.52%
	Plant Height	4	2	27.18%
	Maturity	7	2	49.95%
	100-Seed Weight	4	3	77.46%

	Traits	No.of initial markers	No. of markers included after stepwise regression	R^2
Durango	Seed Yield	1	1	08.72%
	Plant Height	4	1	04.02%
	Maturity	3	2	59.52%
	100-Seed Weight	7	2	13.41%

	Traits	No.of initial markers	No. of markers included after stepwise regression	R^2
Mesoamerican	Seed Yield	73	7	78.69%
	Plant Height	1	1	14.55%
	Maturity	66	7	78.45%
	100-Seed Weight	5	2	23.78%

To determine the total phenotypic variance explained by each marker, R- square values were used. In all populations, for seed yield, R-square values ranged from 0.25% to 42%. For

maturity, the phenotypic variance from 1.65% to 50.35% was observed. R- Square value for plant height ranged from 1.14% to 20.67% and for 100-seed weight it was 6.35% to 70.06%. For seed yield and maturity highest R- square value was found in Mesoamerica population while for seed weight and plant height, the highest R- square value was seen in AYT + PYT population (Appendix A).

In this study many of the significant markers associated with a particular trait were observed in more than one population (Appendix A). For seed yield, marker sc00853ln138233_81612_G_A_292173871 at 39.3 Mbp on chromosome 7 was found in populations AYT+PYT, AYT and Mesoamerica and marker sc00013ln1423374_1157977_A_C_22673519 at 42.38 Mbp on chromosome 7 was found in populations AYT+PYT and Durango. For 100-seed weight and plant height, only one marker was observed which was common in more than one population. Marker sc00055ln737569_432814_C_T_61008497 at 29.79 Mbp on chromosome 9 for seed wright was found in both AYT+PYT and AYT populations with highest R- square value 41.39%. Marker sc00296ln326650_106196_C_A_175840672 at 52.73 Mbp on chromosome 8 which was found to be significantly associated with plant height was observed in PYT and AYT+PYT populations.

For maturity, marker sc00853ln138233_81612_G_A_292173871 at 39.3 Mbp on chromosome 7 was found in AYT+PYT and Mesoamerica, marker sc00055ln737569_432814_C_T_61008497 at 29.79 Mbp on chromosome 9 was found in AYT+PYT and AYT.

Some significant markers were found to be associated with more than one trait. Marker sc00853ln138233_81612_G_A_292173871 was significantly associated with both seed yield and

maturity in population AYT+PYT and Mesoamerica. For seed weight and maturity, another marker, sc00055ln737569_432814_C_T_61008497 on chromosome 9 was identified in populations AYT and AYT + PYT.

QTL regions were identified for seed yield and maturity based on the markers which were found in linkage disequilibrium with other markers. For both yield and maturity, marker sc00701ln164739_86002_G_A_269264751 at 38.82 Mbp, on chromosome 7 was found to be in LD with sc00853ln138233_81612_G_A_292173871 at 39.29 Mbp indicating a QTL region. For maturity, another QTL region was found. QTL region spanning 0.29 Mbp for maturity was detected with three flanking markers sc00015ln1350335_1191524_C_T_25527800 at 36.68 Mbp, sc00015ln1350335_1210315_C_T_25546591 at 36.69 Mbp and sc00015ln1350335_1224361_T_G_25560637 at 36.70 Mbp on chromosome 9. Unlike seed yield and days of maturity, no QTL was observed for plant height and 100-seed weight as no markers was found in LD with either of the two markers (Table 8).

Table 8. QTL and the candidate genes with their positions.

Trait	QTL	Flanking Markers	Ch	Physical position (Mbp)	Gene model	Candidate gene	Gene Physical position (Mbp)
Seed Yield	SYMAT7	sc00853ln138233_81612_G_A	7	39.2	Phvul.007G162100	IPT	39.04
		sc00701ln164739_86002_G_A	7	38.8	Phvul.007G161600	GTP-binding protein 1	38.98
Maturity	SYMAT7	sc00853ln138233_81612_G_A	7	39.3	Phvul.009G254300	Ribosomal protein S26e family protein	36.69
		sc00701ln164739_86002_G_A	7	38.8	Phvul.007G161600	GTP-binding protein 1	38.98
Maturity	MAT9	sc00015ln1350335_1191524_C_T	9	36.68	Phvul.009G254200	WD-40 repeat family protein	36.68
		sc00015ln1350335_1224361_T_G	9	36.71	Phvul.009G254400	salt-inducible zinc finger 1	36.71

Allelic combinations

Different allelic combinations associated with the traits were identified in all subpopulations (Table 8). For maturity, a genotype with less than 90 days was considered as

early maturing, and genotypes maturing after 112 days as late maturing and rest as average. According to this, in population AYT, two allelic combinations (AAG and GAG) could be associated with early maturity. In Mesoamerican subpopulation, a total of 38 allelic combinations were found for early maturity. In Durango, all allelic combinations were associated with average days to maturity while in population AYT+PYT, all combinations obtained could be responsible for early maturity.

For seed yield, the range for high and low seed yield decided was 3040 and 2159 kg ha^{-1}, respectively. Allelic combinations contributing to high yield were found only in Mesoamerica (5 combinations) and PYT (1 combination) subpopulations. In the rest of sub-populations allelic combinations might be contributing to an average seed yield.

For plant height and 100-seed weight, cut off values were decided separately for the landraces as both landraces have different characteristic traits. For Durango, plant height more than 60 cm were considered as tall and plants less than 50 cm as short. For Mesoamerica sub-population, 50 – 60 cm was decided as average range for plant height. For sub-populations AYT, PYT, and AYT+PYT the range for short and tall plant was 50 cm and 59 cm respectively. Based on these cut off values, all combinations in Durango and Mesoamerica sub-populations were thought to be responsible for average plant height. In AYT and PYT sub-populations, only one allelic combination contributed to taller plants and the remaining combinations with average plant height. For AYT+PYT sub-population, 2 allelic combinations contributed to plants with taller height and 2 combinations were found to be associated with short plants.

Similarly for 100-seed weight, the cutoff values were 21g - 40g for AYT, PYT and AYT+PYT, 36g – 42g for Durango and 19g - 22g were for Mesoamerica populations. Allelic combination, obtained for genotypes of AYT, PYT, AYT+PYT and Mesoamerica were found to

be associated with an average seed weight. While in Durango, one combination was found that could be responsible for higher seed weight.

Table 9. Allelic combinations observed in all markers for all agronomic traits in AYT.

Trait (AYT)	Combinations	No. Obs.	Minimum	Maximum	Mean	Std. Dev.	
Maturity	AAG	1	75	75	75	.	Early
(days)	GAG	37	76	104	80	6	Early
	GAA	15	77	108	96	14	Average
	AGA	12	78	106	102	8	Average
	AGG	7	82	110	102	9	Average
	GGG	11	77	113	103	13	Average
	GGA	48	79	111	105	5	Average
	AAA	5	102	108	106	2	Average
Seed Yield	AA	22	1657	2662	2109	288	Low
(kg ha-1)	AG	7	1814	2509	2343	248	High
	GA	44	1960	2927	2378	231	High
	GG	63	1939	3390	2779	350	High
100-Seed	AAA	6	19	22.7	20.9	1.4	Low
Weight (g)	GAG	6	18.8	20.5	19.8	0.6	Low
	AAG	44	17.8	24.5	20.8	1.5	Low
	AGG	18	19.1	24.3	21.7	1.4	Low
	GAA	2	24.4	41.1	32.8	11.8	Average
	AGA	32	19	44.8	33.6	6.8	Average
	GGA	28	19	41.9	35.7	5.5	Average
Plant	AA	11	50	62	56	4	Average
Height	AC	106	47	63	54	3	Average
(cm)	GC	9	53	62	57	3	Average
	GA	10	53	64	60	3	Tall

The marker order that was selected to determine allelic combination were in a sequence of:
Maturity: sc00789ln147969_135375_A_G, sc00055ln737569_432814_C_T, sc00015ln1350335_1191524_C_T.
Seed yield: sc00675ln170111_135617_A_G, sc00853ln138233_81612_G_A.
100-seed weight: sc00293ln329559_110683_C_T, sc00055ln737569_432814_C_T, sc00015ln1350335_515184_A_G.
Plant height: sc00268ln345453_135128_C_T, sc08776ln2265_1282_A_C.

Table 10. Allelic combinations observed in all markers for all agronomic traits in Durango.

Trait (Durango)	Combinations	No. Obs.	Minimum	Maximum	Mean	Std. Dev	
Maturity (days)	AA	45	106	111	109	4	Average
	GA	8	113	113	113	0	Late
	AG	71	101	108	105	5	Average
	GG	2	108	110	109	1	Average
Seed Yield ($kg\ ha^{-1}$)	A	45	1807	3423	2652	376	High
	C	81	1394	3565	2397	410	High
100-Seed Weight (g)	AA	45	27.2	44.8	35.4	4.9	Low
	AG	6	22.9	39.7	31.8	6.8	Low
	GG	4	29.1	34.2	32.2	2.4	Low
	GA	71	24.9	43.1	37.4	3.2	Low
Plant Height (cm)	A	103	41	66	56	4	Average
	G	23	52	69	58	4	Average

The marker order that was selected to determine allelic combination were in a sequence of:
Maturity: sc00382ln273856_226597_C_T, sc00186ln436341_356682_A_G.
Seed yield: sc00013ln1423374_1157977_A_C.
100-seed weight: sc00038ln842375_617118_G_A, sc00324ln307318_62459_C_T
Plant height: sc00384ln271106_234031_C_T.

Table 11. Allelic combinations observed in all markers for all agronomic traits in PYT.

Trait (PYT)	Combinations	No. Obs.	Minimum	Maximum	Mean	Std. Dev	
Seed Yield (kg ha$^{-1)}$	GGCA	2	1394	1714	1554	226	Low
	GACA	2	1467	2078	1772	432	Low
	GGAG	2	2033	2167	2100	95	Low
	GAAA	10	1865	2598	2255	222	Average
	GGAA	3	2234	2595	2468	203	Average
	AGAA	17	2177	3017	2532	199	Average
	GAAG	2	2542	2699	2621	111	Average
	AAAA	19	1837	3565	2704	342	Average
	AGAG	2	2669	2755	2712	61	Average
	AAAG	13	2684	3423	3120	255	High
100-Seed Weight (g)	A	61	24.9	44.7	36	4.6	Average
	G	11	32.5	42	37.3	3.1	Average
Plant Height (cm)	ACG	3	41	52	48	6	Short
	GCG	2	50	59	54	6	Average
	AAA	2	56	56	56	0	Average
	ACA	21	52	65	56	4	Average
	GCA	22	53	63	58	3	Average
	GAA	22	55	69	60	4	Tall

The marker order that was selected to determine allelic combination were in a sequence of:
Seed yield: sc02995ln26285_839_A_G, sc00090ln635406_393133_G_A, sc00358ln289292_188590_G_T, sc00014ln1397360_651244_G_A.
100-seed weight: sc00T91ln330179_T5T411_G_A.
Plant height: sc00387ln270001_24153_C_T, sc00296ln326650_106196_C_A, sc00187ln435150_46434_T_C.

Table 12. Allelic combinations observed in all markers for all agronomic traits in AYT+PYT.

Trait (AYT+PYT)	Combinations	No. Obs.	Minimum	Maximum	Mean	Std. Dev	
Maturity (days)	AC	86	77	113	95	25	Early
	CC	60	77	113	95	25	Early
	AA	31	75	107	91	22	Early
	CA	31	77	108	92	22	Late
Seed Yield (kg ha^{-1})	AG	71	1394	3181	2316	363	Average
	CG	34	1467	3565	2547	348	Average
	AA	46	1807	3423	2608	372	Average
	CA	57	1939	3390	2771	387	Average
100-Seed Weight (g)	ACA	41	17.8	24.5	20.7	1.6	Low
	CCA	18	18.9	38.8	22.1	4.5	Average
	ACC	18	19.1	32.7	23.2	4.1	Average
	CCC	30	19.0	39.7	28.6	5.8	Average
	CAA	3	27.2	41.1	34.9	7.1	Average
	AAC	4	32.5	38.1	36.4	2.6	Average
	CAC	94	22.9	44.8	37.5	3.7	Average
Plant Height (cm)	GA	138	41	65	54	3	Average
	GC	35	51	65	58	3	Average
	AC	25	53	66	60	3	Tall
	AA	10	56	69	60	4	Tall

The marker order that was selected to determine allelic combination were in a sequence of:
Maturity: sc00853ln138233_81612_G_A, sc00055ln737569_432814_C_T.
Seed yield: sc00853ln138233_81612_G_A, sc00013ln1423374_1157977_A_C.
100-seed weight: sc00835ln140787_33631_T_C, sc01014ln116695_42911_C_T, sc00055ln737569_432814_C_T.
Plant height: sc00296ln326650_106196_C_A, sc00211ln404231_317008_A_G.

Table 13. Allelic combinations observed in all markers for all agronomic traits in Mesoamerican.

Trait (Mesoamerican)	Combinations	No. Obs.	Minimum	Maximum	Mean	Std. Dev	
Seed Yield	GCGGAAG	1	2082	2082	2082	.	Low
(kg $ha^{-1)}$	AAAAGAG	1	2091	2091	2091	.	Low
	GCAGACG	1	2115	2115	2115	.	Low
	GCAGGAG	3	1943	2361	2132	212	Low
	GAAGACA	1	2185	2185	2185	.	Average
	GAAGGAG	2	1939	2496	2218	394	Average
	GCAGAAG	8	2033	2721	2270	211	Average
	GCGGAAA	2	2284	2440	2362	110	Average
	AAAGACA	1	2364	2364	2364	.	Average
	GCAGAAA	14	2068	2807	2381	198	Average
	GCGGGAG	3	2230	2578	2425	177	Average
	GAAGGCA	3	2232	2804	2520	286	Average
	GCAGGAA	1	2586	2586	2586	.	Average
	ACAGGAA	6	2363	2754	2607	157	Average
	GCAGACA	1	2623	2623	2623	.	Average
	GCGGGAA	4	2268	3012	2726	327	Average
	ACGGGCA	1	2920	2920	2920	.	Average
	GCAGGCA	2	2824	3022	2923	139	Average
	GCGGGCA	4	2664	3223	2959	230	Average
	GCAAGCA	3	2761	3337	2970	319	Average
	GAGAGCA	2	2878	3098	2988	156	Average
	ACGAGCA	7	2852	3190	3023	100	Average
	GCGAGAA	2	2720	3390	3055	474	High
	ACAAGCA	3	2914	3216	3071	152	High
	GCGAGCA	4	2912	3372	3078	201	High
	ACGAGAA	1	3322	3322	3322	.	High
	GAAAGCA	1	3362	3362	3362	.	High
100-Seed	AA	53	17.8	24.5	20.5	1.4	Average
Weight (g)	AC	6	20.2	22.7	21.5	1	High
	GA	15	19.1	24.3	21.5	1.6	High
	GC	8	19.6	25.9	23.1	1.9	High
Plant Height	A	6	47	55	51	3	Average
(cm)	G	76	48	63	55	3	Average

The marker order that was selected to determine allelic combination were in a sequence of:
Seed yield: sc00174ln464616_175582_G_A, sc01016ln116177_107973_T_G, sc00779ln149779_105659_A_G, sc02498ln35994_11115_G_A, sc00853ln138233_81612_G_A, sc00015ln1350335_1224361_T_G, sc00119ln552178_203683_C_T.
100-seed weight: sc00327ln306820_257821_C_T, sc00071ln681296_466314_A_C.
Plant height: sc01058ln111657_20684_G_A.

Table 13. Allelic combinations observed in all markers for all agronomic traits in Mesoamerican. (continued)

Trait (Mesoamerican)	Combinations	No. Obs.	Minimum	Maximum	Mean	Std. Dev	
Maturity (days)	AGGGAGA	1	75	75	75	.	Late
	AGAGGAA	1	77	77	77	.	Late
	CAAAGAA	1	77	77	77	.	Late
	CGAGAGA	1	77	77	77	.	Late
	AAGGAGA	1	77	77	77	.	Late
	AGGGGAA	1	77	77	77	.	Late
	AAAGAAC	1	78	78	78	.	Late
	CGAGGGC	1	78	78	78	.	Late
	AAAGAAA	1	78	78	78	.	Late
	AAAGAGA	19	76	80	78	1	Late
	CGAGAGC	2	77	79	78	1	Late
	AAAGGAA	1	78	78	78	.	Late
	AGAGAGC	1	78	78	78	.	Late
	AGAGGGC	1	79	79	79	.	Late
	CAAGAGA	1	79	79	79	.	Late
	AGAGGGA	1	79	79	79	.	Late
	AAAGGGA	7	79	82	80	1	Late
	CGGGAGA	1	82	82	82	.	Late
	CGAGGAA	2	78	104	91	18	Late
	CAGGGGA	4	77	106	92	15	Average
	AGGAGAA	1	102	102	102	.	Average
	CGGGGAA	1	102	102	102	.	Average
	CGAAGAC	2	103	104	103	1	Average
	CAGGGAA	1	104	104	104	.	Average
	CGGAGAC	7	103	107	104	1	Average
	CAAAGAC	3	103	107	105	2	Average
	CGAGGAC	2	104	106	105	2	Average
	AAGAGAC	1	105	105	105	.	Average
	CGGGGAC	1	105	105	105	.	Average
	CAGAGAC	5	105	106	106	1	Average
	CAGGGAC	4	105	108	106	1	Average
	AAAAGGC	1	107	107	107	.	Average
	CAGAGAA	2	107	107	107	0	Average
	CAAAGGC	1	107	107	107	.	Average
	CAAGGAC	1	108	108	108	.	Average

The marker order that was selected to determine allelic combination were in a sequence of: Maturity: sc00893ln132397_78914_G_T, sc00183ln440474_149829_C_T, sc00779ln149779_105659_A_G, sc02498ln35994_11115_G_A, sc00853ln138233_81612_G_A, sc00262ln351368_226896_C_T, sc00015ln1350335_1224361_T_G.

DISCUSSION

Genome wide association mapping (GWAS) is one of several approaches developed to identify marker-trait associations taking advantage of linkage disequilibrium. This approach was initially used in human and animals and has been adapted in plants for different purposes. In this study, GWAS is used to identify statistical association between markers across the whole genome and quantitative traits of agronomic/economic interest using genotypes from a breeding program.

The population used in this experiment consisted of advanced breeding lines and preliminary breeding lines from the NDSU dry bean breeding program. These are the superior lines that are selected and developed after multiple testing and re-evaluations. These breeding lines involve collection of genotypes from different parents that represents the genetic diversity within the breeding program. Using these lines from the breeding program as a population for GWAS, instead of using diversity panels is a new practical approach for the development of cultivars which also proves that breeding lines can be used for genome wide studies and QTL studies as mapping populations and can save money. These lines contain higher genetic diversity and high mapping resolution that will allow to select more efficient markers associated with breeder-targeted traits, and to use them directly in MAS once QTLs are identified in much less time. It can give more robust results for the maker- agronomic trait associations in the superior lines currently being selected and developed in North Dakota. In this way, the breeding populations can be used for cultivar development and genetic studies simultaneously.

GWAS was conducted, combining AYT and PYT together and separately as PYT, AYT, and based on races Mesoamerican and Durango using polymorphic SNP markers. The breeding lines were sub-divided into groups to cross check whether or not significant markers discovered

in one sub-population are also significantly associated with the same trait in another subpopulation and to have better understanding about genetic diversity and relatedness among varieties of beans within both Mesoamerican races.

Several studies have been made in the recent years to find significant markers and candidate genes associated with seed yield, maturity, seed weight, and plant height by using QTL mapping utilizing different marker systems. Seven populations consisting of Mesoamerican gene pools were used in QTL mapping, for seed weight, using seed protein-based markers. *Phs* locus was found to be associated with 100-seed weight at $p < 0.001$ in an experiment conducted by Bitocchi *et al*. (2011).

In this analysis, a QTL region associated with days to maturity, identified on the chromosome 7 and chromosome 9 may be similar to the ones observed in earlier studies. For example, in a study conducted to identify genetic loci associated with 14 quantitative traits in common bean using RAPD, SSR, and AFLP markers, found the nearest locus for days to maturity and seed yield on chromosome 9 (Tar'an *et al*., 2002). Another QTL region for days to maturity was identified from 96 RILs from a 'Jaguar'/115M black bean cross, using SSR, SRAP, SCAR and TRAP was identified on chromosome 7 (Wright *et al*., 2011). Significant association of SSR markers with maturity and seed weight on chromosome 7 and 8 was detected in a study conducted with Bat93 x Jalo EEP558 population (Reinprecht *et al.,* 2012). Few significant markers associated with days to maturity, yield, plant height, and 100-seed weight were identified on chromosome 7, chromosome 8, and chromosome 9 using a diverse panel of a Mesoamerican dry bean population and 10k SNP chip (Moghaddam, personal communication). 13 QTL were found using SCAR markers associated with seed weight, plant height, and seed yield in 157 $BC_2F_{3:5}$ introgression lines generated from the cross between Colombian large red-

seeded commercial cultivar ICA-Cerinza and a wild common bean accession, G24404 (Blair *et al*., 2006). In earlier studies, most of the QTL regions were already identified on chromosomes 7 and chromosome 9, this analysis found two new QTL region for both traits i.e. seed yield and days to maturity on same chromosomes. In addition to this, breeding-specific significant markers (for all four traits) and candidate genes (for days to maturity and seed yield) within the QTL region were identified in this study.

In the present GW Association Study, many genetic components were discovered which could be responsible and involved in improving these agronomic traits of dry beans. Some of the significant markers that were identified in GWAS were associated with more than one trait of interest and were present in more than one subpopulation, indicating the consistency of the marker- trait association. Across all populations, 2 new QTL and 13 significant markers for seed yield, 11 significant markers for maturity, 9 significant markers for 100-seed weight, and 9 significant markers for plant height were detected. These newly identified markers, QTL, candidate genes within these QTL can be directly used in MAS.

Candidate genes

Candidate gene analysis was done within estimated QTL identified for seed yield and days to maturity. This analysis was done by targeting the genes with already known functions. This analysis was proven successful in pine (*Pinus* spp.) (Gonzalez-Martinez *et al*. 2007) and maize (Wilson *et al*., 2004; Weber *et al*., 2007), where genes responsible for variation in wild and cultivated species were identified.

There are three stages that are involved in seed development. These are embryogenesis, endosperm development and seed maturation in which cell division starts in the embryo. During seed maturation it undergo embryo growth, seed filling, desiccation phase, and at end to

quiescent state. During desiccation phase, dormancy is included which indicate seed germination time. Seed maturation is a process that starts from ovule fertilization and after many morphological, physical, physiological and biochemical changes develops into an independent seed (Miller *et al.*, 1999). Studies showed that seed coat is an important factor that influences embryo growth by interfering water uptake, gas exchange, and diffusion of endogenous inhibitors (Watkins *et al.*, 1983). Similarly, many factors are responsible for increase and decrease in seed yield. It can be due to environment effect, diseases, biotic and abiotic stress or can be caused by biochemical changes. These previous studies suggest that the genes identified in this research could be responsible for either maturity or yield in dry beans.

A total of six genes were found to be associated with seed yield and days to maturity. For maturity, two candidate genes were found within the QTL region that was flanked by three markers sc00015ln1350335_1210315_C_T_25546591, sc00015ln1350335_1191524_C_T_25527800 and sc00015ln1350335_1224361_T_G_25560637 on chromosome 9. The first gene estimated within this QTL region, is a transducin family protein / WD-40 repeat family protein. WD-40 proteins are found in both animals and plants representing a large family in eukaryotes. A highest level of this single copy gene (lacking introns) is expressed in seed coats. WD-40 encodes open reading frames (343 amino acids) (Pang *et al*., 2009) and consists of 4 to 16 repeating units, which assembles to form a circularized beta propeller structure. These repeating units comprise for 40 amino acids core which ends with tryptophan- aspartic acid (WD) (Smith *et al*., 1999, Ramsey *et al*., 2005, Ito *et al*., 2001). In *Arabidopsis*, GIGANTUS1 (GTS1), a member of WD-40 protein which was highly expressed during seed germination by interacting with ribosomal proteins to regulate cell growth (Gachomo *et al*., 2014). This gene is found mainly in seed embryo, ovule, and endosperm. In flax or linseed

(*Linum usitatissimum L.)*, LuWD40-1 genes which encodes WD-40 protein was expressed during vegetative stages found responsible for the regulation of growth and pollen viability (Kumar *et al*., 2013). WD-40 protein was also identified in metabolic pathway and developmental stages of sugar beet (*Beta vulgaris L.*) (Bellin *et al*., 2007) and *rice* (Huang *et al.*, 2008). Expression of MtWD-40 gene was also observed in Barrel Clover (*Medicago truncatula L.)* during seed development. It was found that shortage of this can block accumulation of many compounds such as mucilage, phenolic and flavonoids in the seed. (Pang *et al*., 2009).

Second candidate gene is a salt-inducible zinc finger 1 gene which could be associated with maturity. NFX1-type is one of several zinc protein domain that expresses MHC II gene, is involved in many aspects of growth and development of many crops by managing salicylic acid, reactive oxygen species, and abscisic acid responses under abiotic and biotic stress conditions (Ciftci-Yilmaz *et al*., 2007). Gene GhZFP1 in zinc finger protein 1 in cotton OsOSAPI gene (isolated from rice) in transgenic tobacco played central role in stress signaling (Gua *et al*., 2009) and in tomato it conferred salinity tolerance (Mukhopadhyay *et al*., 2004). Zinc finger proteins are found in eukaryotes and act as transcription factors in many other plants. Motifs of zinc finger protein are found in proteins are known as TFIIIA, which can bind to DNA through amino acid interaction of DNA base pair with zinc finger (Takatsuji *et al*., 1988). This DNA binding domain has leucine rich repeats sites that determine the regulatory function in stress conditions. Based on structure and functions of zinc finger protein is found in different types such as C2H2, C8, C6, C3HC4, C2HC, C2HC5, C4, C4HC3 and CCCH where H is histidine and C is cysteine (Miller *et al*., 1985).

For both seed yield and maturity, few genes were identified on the same QTL that was flanked between markers sc00853ln138233_81612_G_A_292173871 and

sc00701ln164739_86002_G_A_269264751 on chromosome 7. Gene ribosomal protein S15a is involved as a growth regulator in *Arabidopsis* which is divided into type I and type II. Type I is a ribosomal cytosolic component whereas type II position is unclear between cytosolic and/or mitochondrial ribosomes. In 2009, Szick-Miranda investigated that type II gene is a regulator for translational activity. Significant difference of ribosomal protein S15a regulation was observed in seed maturity, seed embryo and desiccation between two cultivars of rapeseed (*Brassica napus L.*) (AC Excel and DH12075) (Fei *et al.,* 2006). The results suggested that expression of this gene changes during secondary dormancy and thus are involved in seed maturation.

Isopentenyltransferase (IPT) gene a key enzyme in the cytokinin biosynthesis is identified for seed yield and maturity on significant QTL. In earlier studies, expression of this protein in peanut (*Arachis hypogaea L.*), rice, and tobacco (*Nicotiana tabacum L.*) displayed increase in photosynthetic rates, stomatal conductance, and transpiration under drought conditions resulting in a seed yield increase (Hua *et al*., 2011). Another research done on salinized tomato (*Solanum lycopersicum L.*) showed 30% increase in fruit yield due to that the involvement of IPT gene. Transport of cytokinin from root to shoot maintain stomatal conductance, thereby, delay accumulation of toxic Na+ ions and increasing fruit yield (Michell *et al*., 1984).

Gene AT5G52210 encodes a GTP binding protein which play important role in signal transduction and cell differentiation is estimated for seed yield and maturity. A study conducted in maize plants, found that this gene contributes to, many agronomic traits including seed yield (Kang *et al*., 1995). Hydroxyproline-rich glycoprotein family protein (HRGP), identified as a candidate gene in this study also showed its contribution in many yield traits in other plants. This protein supports the cellular organelles of plant by non-covalent interaction between protein

chains. A chain extension is accumulated during pathogen attack as a self-defense (Michael, 1994). The role HRGP in plant resistance was studied previously in melon and beans by Connell *et al*., in 1990. In maize, expression of this gene is found in immature embryos, ovaries and nonvascular cell (Josè-Estanyol *et al*., 1992).

CONCLUSIONS

The study shows that sub-populations from a breeding program can also be used for genome wide studies simultaneously with cultivar development. Using these preliminary and advanced breeding lines from these breeding populations we were able to identify many significant markers and possible candidate genes that could be associated with the agronomic traits studied.

GWAS with five different subpopulations confirmed that markers discovered were significantly associated with the same trait in another subpopulation. After stepwise regression, a considerable set of 13 SNP markers linked to seed yield, 11 SNPs for maturity, 9 SNPs for 100-seed weight and 9 SNPs for plant height have been identified. One significant SNP marker named sc00853ln138233_81612_G_A_292173871 on chromosome 7 was found to be associated with both seed yield and maturity. Another marker sc00055ln737569_432814_C_T_61008497, on chromosome 9 was identified to be associated with both 100- seed weight and maturity. Similarly, 2 SNP markers on chromosome 7 associated with seed yield, 2 SNP markers on chromosome 7 and chromosome 9 for maturity, 1 SNP markers on chromosome 8 for plant height and 1 SNP on chromosome 9 for 100- seed weight were commonly observed in more than one subpopulations increasing the likelihood of the marker trait association (Appendix 1). SNP markers that are identified to be significantly associated with more than one trait or found in more than one subpopulation, could be used directly in MAS for a particular race, AYT or PYT.

Based on the PV annotation data, GWAS identified QTL – SYMAT7 for two agronomic traits, seed yield and maturity, on chromosome 7 and also QTL- MAT9 for days to maturity on chromosome 9 using breeding populations. Within these QTL, ribosomal protein S26e family protein, GTP-binding protein 1, WD-40 repeat family protein, and salt-inducible zinc finger 1

were identified as candidate genes for days to maturity. For seed yield, IPT and GTP-binding protein 1 were identified as candidate genes. A high throughput study (e.g. RNA-Seq) on the expression of these genes could help the breeders to have a better understanding of the role of these genes into these traits of economic importance.

Few of the allelic combinations identified in this study represented that these combinations have significant difference on the extreme of phenotypes.

All of these genetic components (markers, candidate genes, and allelic combinations) identified could now be used to understand genetic basis of these traits by exploring their biological pathway that is expected to be responsible or related for a particular trait. Later these genes can be also used to find synteny with other related crops, or can be transferred to different crops to manipulate the trait of interest. Thus using lines from breeding populations for GWAS and estimated significant markers and candidate genes associated with trait of interest could help to develop better progenies in more efficient way and in short time period.

REFERENCES

Abdurakhmonov, I.Y., and A. Abdukarimov. 2008. Application of association mapping to understanding the genetic diversity of plant germplasm resources. Int. J. Plant Genomics. 2008: 1-18. DOI:10.1155/2008/574927.

Adam-Blondon, A.F., M. Sevignac, M. Dron, and H. Bannerot. 1994. A genetic map of common bean to localize specific resistance genes against anthracnose. Genome 37:915-924.

Atella, D. 2007. Variability, heritability and correlation studies in bean. World J.Agric. Sci. 3: 12-16.

Azarpazhooh, E., and J.I. Boye. 2012. Composition of processed dry beans and pulses. *in* Siddiq, M. and Uebersax, M.A. (eds.) - Dry Beans and Pulses: Production, Processing and Nutrition. John Wiley & Sons, Ltd, Chapter 5, pp. 103-128.

Balding, D.J. 2006. A tutorial on statistical methods for population association studies. Nature Rev. Genet. 7: 781-791.

Becerra-Velásquez, V.L., and P. Gepts. 1994. RFLP diversity in common bean (*Phaseolus vulgaris L.*). Genome 37:256–263.

Beebe, S., P.W. Skroch, J. Tohme, M.C. Duque, F. Pedraza, and J. Nienhuis. 2001. Structure of genetic diversity among common bean landraces of Middle American origin based on correspondence analysis of RAPD. Crop Sci 40:264-273.

Beer, S., W. Siripoonwiwat, L. O'donoughue, E. Souza, D. Mathews and M. Sorrels. 1997. Associations between molecular markers and quantitative traits in oat germplasm pool: can we infer linkages? J. Agric. Genomics, 3:197.

Bitocchi, E., L. Nanni, E. Bellucci, M. Rossi, A. Giardini, P.S. Zeuli, G. Logozzo, J. Stougaard, P. McClean, G. Attene, and R. Papa. 2011. Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *In*: Jeffrey L. Bennetzen, University of Georgia, Athens, GA (editors). DOI: 10.1073/pnas.1108973109.

Bellin. D., B. Schulz, T.R. Soerensen, F. Salamini, and K. Schneider. 2007. Transcript profiles at different growth stages and tap-root zones identify correlated developmental and metabolic pathways of sugar beet. J. Exp. Bot.58:699–715.DOI:10.1093/jxb/erl245.

Blair, M.W., G. Iriarte, and S. Beebe. 2006. QTL analysis of yield traits in an advanced backcross population derived from a cultivated Andean × wild common bean (*Phaseolus vulgaris L.*) cross. Theor. Appl. Genet. 112:1149-1163.

Britten, R.J., L. Rowen, J. Williams, and R.A. Cameron. 2003. Majority of divergence between closely related DNA samples is due to indels. PNAS 100:4661-4665.

Broughton, W.J., G. Hernandez, M. Blair, S. Beebe, P. Gepts, J. Vanderleyden. 2003. Beans (*Phaseolus* spp) – model food legumes. Plant and Soil 252:55-128.

Carsten, F.D., J. Elith, S. Bacher, C. Buchmann, G. Carl, B. Gruber, B. Lafourcade, P.J. Leitão, T. Münkemüller, C. McClean, P.E. Osborne, B. Reineking, B. Schröder, A.K. Skidmore, D. Zurell, and S. Lautenbach. 2012. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 35: 1–20.

Chengsong, Z., M. Gore, E.S. Buckler, and J. Yu. 2008. Status and prospects of association mapping in plants. The Plant Genome 1:5-20.

Ciftci-Yilmaz S., M.R. Morsy, L. Song, A. Coutu, B.A. Krizek, M.W. Lewis, D. Warren, J. Cushman, E.L. Connolly and R. Mittler. 2007. The ear-motif of the C2H2 zinc-finger protein ZAT7 plays a key role in the defense response of *Arabidopsis* to salinity stress. J. Biol. Chem. 282: 9260–9268.

Chacón, S.M.I., B. Pickersgill, and D.G. Debouck. 2005. Domestication patterns in common bean (*Phaseolus vulgaris* L.) and the origin of the Mesoamerican and Andean cultivated races. Theor. Appl. Genet. 110:432–444.

Cortes, A.J., M.C. Chavarro, and M.W. Blair. 2011. SNP marker diversity in common bean (*Phaseolus vulgaris* L.). Theor. Appl. Genet. 123:827–845.

Cuenca, J., P. Aleza, L. Navarro, and P. Ollitrault. 2013. Assignment of SNP allelic configuration in polyploids using competitive allele-specific PCR: application to citrus triploid progeny. Ann Bot. 1-12.

Debouck, D.G., O. Toro, O.M. Paredes, W.C. Johnson, P. Gepts. 1993. Genetic diversity and ecological distribution of *Phaseolus vulgaris* in northwestern South America. Econ. Bot. 47: 408-423.

Delgado-Salinas, A., T. Turley, A. Richman, and M. Lavin. 1999. Phylogenetic analysis of the cultivated and wild species of *Phaseolus* (Fabaceae). Syst. Bot. 24:438-460.

Delgado-Salinas, A., R. Bibler, and M. Lavin. 2006. Phylogeny of the genus *Phaseolus* (Leguminosae): a recent diversification in an ancient landscape. Syst. Bot. 31:779-791.

Doyle, J.J., and J.L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 19:11-15.

Doyle, J.J., and J.L. Doyle. 1990. Isolation of plant DNA from fresh tissue. Focus 12:13-15.

Drenkard, E., B.G. Richter, S. Rozen, L.M. Stutius, N.A. Angell, M. Mindrinos, R.J. Cho, P.J. Oefher, R.W. Davis and F.M. Ausubel. 2000. A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in Arabidopsis. Plant Physiol. 124,

1483-1492.

Ersoz, E.S., J. Yu, and E.S. Buckler. 2008. Applications of linkage disequilibrium and association mapping in maize. *In:* Kriz A, and Larkins B, eds. Molecular genetic approaches to maize improvement. DOI: 10.1007/978-3-540-68922-5_13.

Eskridge, K.M. 2003. Field design and the search for quantitative trait loci in plants. Available at: http://www.stat.colostate.edu/graybillconference2003/Abstracts/Eskridge.

Fei, H., E. Tsang, and A.J. Cutler. 2006. Gene expression during seed maturation in Brassica napus in relation to the induction of secondary dormancy. Genomics 89:419–428.

Flint-Garcia, S.A., A. Thuillet, J. Yu, G. Pressoir, S.M. Romero, S.E. Mitchell, J. Doebley, S. Kresovich, M.M. Goodman, and E.S. Buckler. 2005. Maize association population: A high-resolution platform for quantitative trait locus dissection. Plant J. 44:1054–1064.

Freund, R.J. and R.C. Littell. 1986. SAS System for Regression, 1986 Edition, Cary, NC: SAS Institute Inc.

Freyer, R., P.W. Skroch, V. Geffroy, A.F. Adam-Blondon, A. Shirmohamadali, W.C. Johnson, V. Llaca, R.O. Nodari, P.A. Pereira, S.M. Tsai, J. Tohme, M. Dron, J. Nienhuis, C.E. Vallejos and P. Gepts. 1998. Towards an integrated linkage map of common bean 4: development of an RFLP-based linkage map. Theor. Appl. Genet. 97:847-856.

Gachomo, E.W., J.C. Jimenez-Lopez, L.J. Baptiste, and S.O. Kotchoni. 2014. GIGANTUS1 (GTS1), a member of Transducin/WD40 protein superfamily, controls seed germination, growth and biomass accumulation through ribosome-biogenesis protein interactions in Arabidopsis thaliana. BMC Plant Biol. 14: 37.

Galeano, C.H., A.J. Cortés, A.C. Fernández, Á. Soler, N. Franco-Herrera, G. Makunde, J. Vanderleyden, and M.W. Blair. 2012. Gene-based single nucleotide polymorphism markers for genetic and association mapping in common bean. BMC Genetics, 13:48. DOI:10.1186/1471-2156-13-48.

Gepts, P. 1998. Origin and evolution of common bean: Past events and recent trends. HortScience 33:1124-1130.

Gepts, P. 1990. Biochemical evidence bearing on the domestication of Phaseolus beans. Econ. Bot. 44(3S): 28-38.

Gepts, P., and D.G. Debouck. 1991. Origin, domestication, and evolution of the common bean, *Phaseolus vulgaris*. Common Beans: Research for Crop Improvement, eds Voysest O, Van Schoonhoven A (CAB International, Wallingford, Oxon, United Kingdom), pp 7–53.

Gepts, P., T.C. Osborn, K. Rashka and F.A. Bliss. 1986. Phaseolin-protein variability in wild forms and landraces of the common bean (*Phaseolus vulgaris*): Evidence for multiple centers of domestication. Econ. Bot. 40: 451–468.

Ghanem, M.E., A. Albacete, A.C. Smigocki, I. Frébort, H. Pospíšilová, C. Martínez-Andújar, M. Acosta, J. Sánchez-Bravo, S. Lutts, I.C. Dodd, and F. Pérez-Alfocea. 2010. Root-synthesized cytokinins improve shoot growth and fruit yield in salinized tomato (*Solanum lycopersicum* L.) plants. J. Exp. Bot. 62: 125-140. DOI: 10.1093/jxb/erq266.

Giri, J., S. Vij, P.K. Dansana, and A.K. Tyagi. 2011. Rice A20/AN1 zinc-finger containing stress-associated proteins (SAP1/11) and a receptor-like cytoplasmic kinase (OsRLCK253) interact via A20 zinc-finger and confer abiotic stress tolerance in transgenic Arabidopsis plants. New Phytologist 191:721–732. DOI: 10.1111/j.1469-8137.2011.03740.x.

Gonzalez-Martinez, S.C., N.C. Wheeler, E. Ersoz, C. Nelson, and D.B. Neale. 2007. Association genetics in *Pinus taeda* L. I. Wood property traits. Genetics 175: 399–409.

Grisi, M.C.M., M.W. Blair, P. Gepts, C. Brondani, P.A.A. Pereira, and R.P.V. Brondani. 2007. Genetic mapping of a new set of microsatellite markers in a reference common bean (*Phaseolus vulgaris*) population BAT93× Jalo EEP558. Genet. Mol. Res., 6:691-706.

Guo, Y.H., Y. Yu, D. Wang, C. Wu, G. Yang, J.G. Huang, and C.C. Zheng. 2009. GhZFP1, a novel CCCH-type zinc finger protein from cotton, enhances salt stress tolerance and fungal disease resistance in transgenic tobacco by interacting with GZIRD21A and GZIPR5. New Phytologist 183: 62–75. DOI: 10.1111/j.1469-8137.2009.02838.x.

Huang, J., M.M. Wang, Y.M. Bao, S.J. Sun, L.J. Pan, and H.S. Zhang. 2008. SRWD: a novel WD40 protein subfamily regulated by salt stress in rice (*Oryza sativa* L.). Gene 424:71–79.

Hyten, D.L., Q. Song, E.W. Fickus, C.V. Quigley, J.S. Lim, I.Y. Choi, E.Y. Hwang, M. Pastor-Corrales, and P.B. Cregan. 2010. High-throughput SNP discovery and assay development in common bean. BMC Genomics 11:475.

Ito, S., A. Sakai, T. Nomura, Y. Miki, M. Ouchida, J. Sasaki, and K. Shimizu. 2001. A novel WD40 repeat protein, WDC146, highly expressed during spermatogenesis in a stage-specific manner. Biochem bioph res co. 280:656-663.

Jafar, M., R. Aggarwal, R. Buyyarapu, and S. Kumpatla. 2012. SNP markers and their impact on plant breeding. Int. J. Plant Genomics 2012:728398. doi:10.1155/2012/728398.

Jianming, Y., and E.S. Buckler. 2006. Genetic association mapping and genome organization of maize. Plant biotech/Food biotech 17:155-160.

Josè-Estanyol, M., L. Ruiz-Avila, and P. Puigdomhnech. 1992. A maize embryo-specific gene encodes a proline-rich and hydrophobic protein. The Plant Cell, Vol. 4, 413-423.

Kalavacharla, V., Z. Liu, B.C. Meyers, J. Thimmapuram, and K. Melmaiee. 2011. Identification and analysis of common bean (*Phaseolus vulgaris* L.) transcriptomes by massively parallel pyrosequencing. BMC Plant Biology 11:135.

Kami, J., V.B. Velásquez, D.G. Debouck, and P. Gepts. 1995. Identification of presumed ancestral DNA sequences of phaseolin in *Phaseolus vulgaris*. PNAS 92:1101–1104.

Kang, H.G., Y.S. Nos, Y.Y. Chung, M. Costa, K. An and G. An. 1995. Phenotypic alterations of petals and sepals by ectopic expression of a rice MADS box gene in tobacco. Plant Mol. Biol. 29:1-10.

Kelly, J.D. 2010. The Story of Bean Breeding. Michigan State University. Available online at http://bean.css.msu.edu/_pdf/Story_of_Bean_Breeding_in_the_US.pdf.

Koinange, E.M.K., S.P. Singh, and P. Gepts. 1996. Genetic control of the domestication syndrome in common bean. Crop Sci. 36:1037-1045.

Kumar, S., M.C. Jordan, R. Datla, and S. Cloutier. 2013. The LuWD40-1 gene encoding WD repeat protein regulates growth and pollen viability in flax (*Linum Usitatissimum* L.). PLoS ONE 8: e69124. DOI:10.1371/journal.pone.0069124.

Kwak, M., and P. Gepts. 2009. Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris L.*, Fabaceae). Theor. Appl. Genet. 118:979–992.

Kwak, M., D. Velasco, and P. Gepts. 2008. Mapping homologous sequences for determinacy and photoperiod sensitivity in common bean (*Phaseolus vulgaris*). J. Heredity 99: 283-291.

Mamidi, S., S. Chikara, R.J. Goos, D.L. Hyten, D. Annam, S.M. Moghaddam, R.K. Lee, P.B. Cregan, and P.E. McClean. 2012. Genome-wide association analysis identifies candidate genes associated with iron deficiency chlorosis in soybean. The Plant Genome 4:154-164.

Mamidi, S., M. Rossi, S.M. Moghaddam, D. Annam, R. Lee, and P.E. McClean. 2012. Demographic factors shaped diversity in the two gene pools of wild common bean *Phaseolus vulgaris L.* Heredity 110:267–276.

McClean, P., P.Gepts and J.Kamir. 2004. Genomics and genetic diversity in common bean. In: H.Stalker, E.Brumer, and R.Winson eds. Legume crop genomics. AOCS Press.60-82.

McConnell, M., S. Mamidi, R. Lee, S. Chikara, M. Rossi, R. Papa, and P. McClean. 2010. Syntenic relationship among legumes revealed using a gene-based genetic linkage map of common bean (Phaseolus vulgaris). Theor. Appl. Genet. 121: 1103-1116.

Maras, M., and J. Sustar-Vozlic. 2013. Genetic diversity and dissemination pathways of Common Bean in Central Europe. J. Amer. Soc. Hort. Sci. 138:297–305.

Michael, P.W. 1994. The structure and function of proline-rich regions in proteins. Biochem. J. 297:249-260.

Miller, S.S., L.A. Bowman, M. Gijzen, and B.L.A. Miki. 1999. Early development of the seed coat of soybean (Glycine max). Ann. of Bot. 84: 297-304.

Miller, J., A.D. McLachlan, and A. Klug. 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J. 4: 1609-1614.

Mitchell, P.D., D.J. Chalmers, and P.H. Jerie. 1984. The effects of regulated water deficits on pear tree growth, flowering, fruit growth, and yield. J. Amer. Soc. Hort. Sci. 109:604– 606.

Mukhopadhyay, A., S. Vij, and A.K. Tyagi. 2004. Overexpression of a zinc-finger protein gene from rice confers olerance to cold, dehydration, and salt stress in transgenic tobacco. PNAS101:6309-6314.

O'Connell. R.J., I.R. Brown, J.W. Mansfield, J.A. Bailey, D. Mazau, D. Rumeau, and M.T. Esquerre-Tugaye. 1990. Immunocytochemical localization of hydroproline – rich glycoproteins accumulating in melon and bean at sites of resistance to bacteria and fungi. Mol. Plant-Microbe Interactions 2:33-40.

Pacurar, D. I., M.L. Pacurar, N. Street, J.D. Bussell, T.L. Pop, L. Gutierrez, and C. Bellini. 2012. A collection of Indel markers for map-based cloning in seven *Arabidopsis* accessions. J.Exp. Bot. 63**:**2491-2501.

Pang, Y., J.P. Wenger, K. Saathoff, G.J. Peel, J. Wen, D. Huhman, S.N. Allen, Y. Tang, X. Cheng, M. Tadege, P. Ratet, K.S. Mysore, L.W. Sumner, M.D. Marks, and R.A. Dixon. 2009. A WD40 repeat protein from medicago truncatula is necessary for tissue-specific anthocyanin and proanthocyanidin biosynthesis but not for trichome development. Plant Physiol. 151:1114-1129.

Paredes, O.M., and P. Gepts. 1995. Extensive introgression of Middle American germplasm into Chilean common bean cultivars. Genet. Resour. Crop. Ev. 42:29-41.

Podlich, D.W., C.R. Winkler, and M. Cooper. 2004. Mapping as you go: an effective approach for marker-assisted selection of complex traits. Crop Sci. 44:1560–1571.

Qin, H., Q. Gu, J. Zhang, L. Sun, S. Kuppu, Y. Zhang, M. Burow, P. Payton, E. Blumwald, and H. Zhang. 2011. Regulated expression of an isopentenyltransferase Gene (IPT) in peanut significantly improves drought tolerance and increases yield under field conditions. Plant Cell Physiol. 52:1904–1914 DOI:10.1093/pcp/pcr125.

Ramsay, N.A., and B.J. Glover. 2005. MYB-bHLH-WD40 protein complex and the evolution of cellular diversity. Trends in Plant Sci. 10:63-70.

Reinprecht, Y., D. Frey, R. John, A. Bowman, T. Smith, P. Pauls. 2012. Towards Association Mapping of Seed Phenolics In Common Bean. Poster presentation at Ontario Pulse Crop Committee Research Update. Hensall, Ontario.

Rossi, M., E. Bitocchi, E. Bellucci, L. Nanni, D. Rau, G. Attene, and R. Papa. 2009. Linkage disequilibrium and population structure in wild and domesticated populations of *Phaseolus vulgaris L*. Evol. Appl. 2:504–522.

Sanglard, D.A., V.S. Mafra, C.A.G. Ribeiro, L.C. da Silva, E.G. de Barros, J.E.S. Moreira, A.P.S. Mota, R.P. Vianello, and T.L.P.O. Souza. "RUDÁ× AND 277 RILS: a potential new core mapping population for common bean." Embrapa Arroz e Feijão-Artigo em periódico indexado (ALICE). 2013. Annual Report of the Bean Improvement Coop**.** 56:23-24.

Scheet, P., and M. Stephens. 2006. A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. Am. J. Hum. Genet. 78:629–644.

Schmutz, J., P.E. McClean, S. Mamidi, G.A. Wu, S.B. Cannon, J. Grimwood, J. Jenkins, S. Shu, Q. Song, C. Chavarro, M. Torres-Torres, V. Geffroy, S.M. Moghaddam, D. Gao, B. Abernathy, K. Barry, M. Blair, M.A. Brick, M. Chovatia, P. Gepts, D.M. Goodstein, M. Gonzales, U. Hellsten, D.L. Hyten, G. Jia, J.D. Kelly, D. Kudrna, R. Lee, M.M.S. Richard, P.N. Miklas, J.M. Osorno, J. Rodrigues, V. Thareau, C.A. Urrea, M. Wang, Y. Yu, M. Zhang, R.A. Wing, P.B. Cregan, D.S. Rokhsar, and S.A. Jackson. 2014. A reference genome for common bean and genome-wide analysis of dual domestications. Nature Genet. 46:707–713.

Singh, S.P. 1982. A key for indentification of different growth habits of *Phaseolus vulgaris L*. Annual Report of the Bean Improv. Coop. 25:92-95.

Singh, S.P. 2001. Broadening the genetic base of common bean cultivars. Crop Sci. 41:1659-1675.

Singh, S.P. 2005. Common bean (*Phaseolus vulgaris L*.). In: R. J, Singh and P.P. Jauhar, eds. Genetic resources, chromosome engineering, and crop improvement. Grain Legumes 1:31-38. CRC Press, Boca Raton, FL.

Singh, S.P., P. Gepts, and D.G. Debouk. 1991. Races of common bean (*Phaseolus vulgaris L*., Fabaceae). Econ. Bot. 45:379-396.

Singh. S.P., J.A. Gutierrez, A. Molina, C. Urrea, and P. Gepts. 1991. Genetic diversity in cultivated common bean II: marker-based analysis of morphological and agronomic traits. Crop Sci. 31:23-29.

Singh, S.P., R. Nodari, and P. Gepts. 1991. Genetic diversity in cultivated common bean I: allozymes. Crop Sci. 31:19-23.

Skroch, P.W. and J. Nienhuis. 1995. Qualitative and quantitative characterization of RAPD variation among snap bean (*Phaseolus vulgaris*) genotypes. Theor. Appl. Genet. 91:1078–1085.

Smartt, J. 1988.Morphological, physiological and biochemical changes in *Phaseolus* bean under domestication. In: Gepts P, ed. Genetic resources of *Phaseolus* bean. Dordrecht, the Netherlands: Kluwer, 143-161.

Smith, T.F., C. Gaitatzes, K. Saxena, and E.J. Neer. 1999. The WD repeat: a common architecture for diverse functions. Trends in Biochem. Sci. 24:181-185.

Sonnante, G., T. Stockton, R.O. Nodari, V.L. Becerra-Velasquez, P. Gepts. 1994. Evolution of genetic diversity during the domestication of common-bean (*Phaseolus vulgaris L*). Theor. Appl. Genet. 89:629-635.

Souza, T.L.P.O., E.G.D Barros, C.M. Bellato, E.Y. Hwang, P.B. Cregan, and M.A. Pastor-Corrales. 2009**.** Nucleotide polymorphisms (SNPs) in common bean. Annu. Rept. Bean Improv. Coop. 52:6-7.

Szick-Miranda, K., A.S. Zanial, A.S. Zanial, S. Abidayo and K.L.C. Slater. 2006. Analysis of RPS15aE, an isoform of a plant-specific evolutionarily distinct ribosomal protein in *Arabidopsis thaliana*, reveals its potential role as a growth regulator. Plant Mol Biol Rep. DOI 10.1007/s11105-009-0148-6.

Takatsuji, H. 1998. Zinc-finger transcription factors in plants. Cell Mol. Life Sci. 54:582-96.

Tar'an, B., T.H. Michaels and K.P. Pauls. 2002. Genetic mapping of agronomic traits in common bean. Crop Sci. 42:544-556.

Tohme, J., D.O. Gonzalez, S. Beebe and M.C. Duque. 1996. AFLP analysis of gene pools of a wild bean core collection. Crop Sci. 36:1375-1384.

Thoday, J. M. 1961. Location of polygenes. Nature 191: 368–370.

USDA-NASS. 2013. Dry Edible Beans, Statistics by Subject, National Agricultural Statistics Service. http://quickstats.nass.usda.gov/#F0FF6C9C-58DA-3410-90F1-7B9BF3207049.

Vallejos, C.E., N.S. Sakiyama, and C.D. Chase. 1992. A molecular marker-based linkage map of *Phaseolus vulgaris L*. Genetics 131:733-740.

Virk, P., B.V. Ford-Lloyd, M.T. Jackson, and H.J. Newbury. 1996. Predicting quantitative variation within rice germplasm within rice germplasm using molecular markers. Heredity 76:296-304.

Wang, M., N. Jiang, T. Jia, L. Leach, J. Cockram, R. Waugh, L. Ramsay, B. Thomas, and Z. Luo. 2011. Genome-wide association mapping of agronomic and morphologic traits in highly structured populations of barley cultivars. Theor. Appl. Genet. pp. 1-14, DOI:10.1007/s00122-011-1697-2.

Wang, J., P.E. McClean, R. Lee, R.J. Goos, and T. Helms. 2008. Association mapping of iron deficiency chlorosis loci in soybean (Glycine max L. Merr.) advanced breeding lines. Theor. Appl. Genet. DOI 10.1007/s00122-008-0710-x.

Watkins, J.T., and D.J. Cantlife. 1983. Mechanical resistance of the seed coat and endosperm during germination of *Capsicum annuum* at low temperature. Plant Physiol. 72:146-150.

Weber, A.P.M., K.L. Weber, C. Wilkerson, and J.B. Ohlrogge. 2007. Sampling the *Arabidopsis* transcriptome with massively parallel pyrosequencing. Plant Physiol. 144: 32–42.

Wilson, L.M., S.R. Whitt, A.M. Ibanez, T.R. Rocheford, M.M. Goodman, and E.S. Buckler. 2004. Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell 16:2719–2733.

Wright, E.M., and J.D. Kelly. 2011. Mapping QTL for seed yield and canning quality following processing of black bean (*Phaseolus vulgaris L.*). Euphytica 179:471–484.

Yamasakia, M., M.I. Tenaillon, I.V. Bi, S.G. Schroeder, H. Sanchez-Villeda, J.F. Doebley, B.S. Gaut, and M.D. McMullen. 2005. A Large-Scale Screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. The Plant Cell 17:2859-2872.

Yu, Z.H., R.E. Stall, and C.E. Vallejos. 1998. Detection of genes for resistance to common bacterial blight of beans. Crop Sci. 38:1290-1296.

Zhou, X., and M. Stephens. 2012. Genome-wide efficient mixed-model analysis for association studies. Nature Genetics 44:821–824.

Zhu, C., M. Gore, E. Buckler, and J. Yu. 2008. Status and prospects of association mapping in plants. The Plant Genome 1:5-20.

Zuo, Q., J. Hou, B. Zhou, Z. Wen, S. Zhang. 2013. Identification of QTLs for growth period traits in soybean using association analysis and linkage mapping. Plant Breed. 132:317–323. doi:10.1111/pbr.12060.

APPENDIX. SIGNIFICANT MARKER TABLE

Statistical summary of single nucleotide polymorphisms (SNPs) significantly associated with four agronomic traits in five different populations.

Markers (PYT)	Ch	Mbp	Log10P	A1	N	Mean	A2	N	Mean	Diff	R2 (%)	MAF	Step wise	FDR_p
Seed Weight														
sc00291ln330179_252411_G_A_174341808	3	33.56	3.89	A	67	58	G	5	50.5	8	1.24	15.28		0.366
Yield														
sc02995ln26285_839_A_G_423910252	5	2.44	3.97	A	51	2753	G	21	2193	560	34.75	29.17	Yes	0.076
sc00090ln635406_393133_G_A_84898951	5	40.36	3.19	A	46	2680	G	26	2430	250	7.74	36.11	Yes	0.2574
sc00358ln289292_188590_G_T_194953953	6	19.79	5.08	A	68	2644	C	4	1663	981	27.07	5.56	Yes	0.0111
sc00014ln1397360_651244_G_A_23590160	7	0.88	3.37	A	53	2472	G	19	2917	445	20.65	26.39	Yes	0.1985
sc00006ln1798808_286665_A_G_10518130	10	41.02	6.97	A	45	2463	G	27	2801	339	14.42	37.5		0.0003
sc00006ln1798808_212015_C_A_10443480	10	40.95	4.93	A	45	2474	C	27	2783	309	12.01	37.5		0.0111
sc00006ln1798808_183339_A_G_10414804	10	40.92	3.76	A	29	2779	G	43	2462	317	12.97	40.28		0.0974
Plant Height														
sc00387ln270001_24153_C_T_202900033	4	38.82	3.44	A	26	55	G	46	59	4	16.73	36.11	Yes	0.4398
sc00187ln435150_46434_T_C_135198657	8	59.45	4.02	A	67	58	G	5	51	8	19.79	6.94	Yes	0.2657
sc00296ln326650_106196_C_A_175840672	8	52.73	3.28	A	24	60	C	48	56	4	17.23	33.33	Yes	0.4398
sc01406ln80070_59175_G_A_350366930	11	5.36	3.2	A	44	56	G	28	60	3	15.49	38.89		0.4398

Ch. and Mbp is the chromosomes number and position of the significant marker respectively.
$Log_{10}P$ is log10 P- value of the SNP.
A1 and A2 represent allele-1 and allele-2
N is number of alleles observed in that marker.
Diff. is difference between the means of both alleles calculated.
R is the value of R- square represented in percentage.
Stepwise shows all the markers that were included in stepwise regression model.

Significant marker table (continued).

Markers (AYT+PYT)	Ch	Mbp	Log10P	A1	N	Mean	A2	N	Mean	Diff	R2 (%)	MAF	Stepwise	FDR_p
Seed Weight														
sc00835ln140787_33631_T_C_289614938	4	0.51	3.65	G	145	33.7	A	63	22.4	11.3	39.7	30.29	Yes	0.341
sc01014ln116695_42911_C_T_312567327	6	11.58	3.18	A	107	23.6	A	101	37.4	13.8	70.1	48.56	Yes	0.503
sc02958ln26822_20802_C_T_422948602	6	11.46	3.18	G	107	23.6	A	101	37.4	13.8	70.1	48.56		0.503
sc00055ln737569_432814_C_T_61008497	9	29.79	4.09	G	146	33.9	A	62	21.8	12.1	44.9	29.81	Yes	0.245
Yield														
sc00675ln170111_135617_A_G_264966309	2	42.07	3.25	G	160	2580	A	48	2421	158	4.64	23.08		0.433
sc00013ln1423374_1157977_A_C_22673519	7	42.38	4.07	C	105	2576	A	103	2509	67	14.2	49.52	Yes	0.253
sc00853ln138233_81612_G_A_292173871	7	39.3	3.78	A	117	2570	G	91	2508	62	9.73	43.75	Yes	0.253
sc01477ln74590_51381_G_A_355854774	9	0	3.25	A	91	2570	G	117	2522	47	4.82	43.75		0.433
Maturity														
sc00611ln188451_70990_T_C_253445316	4	45.79	3.07	G	196	64	A	12	53	11	2.49	5.77		0.368
sc00002ln2152649_1976120_C_A_4148171	7	51.59	5.36	C	71	79	A	137	55	24	1.29	34.13		0.013
sc00853ln138233_81612_G_A_292173871	7	39.3	4.46	G	91	76	A	117	53	24	5.47	43.75	Yes	0.053
sc00862ln137460_111453_C_T_293443775	8	4.92	3.46	G	25	102	A	183	58	44	3.79	12.02		0.176
sc00862ln137460_62873_C_T_293395195	8	4.87	3.46	A	183	102	G	25	58	44	3.79	12.02		0.176
sc00789ln147969_135375_A_G_283059614	8	4.79	3.46	G	183	58	A	25	102	44	3.79	12.02		0.176
sc00055ln737569_432814_C_T_61008497	9	29.79	4.29	A	146	56	G	62	81	25	44.4	29.81	Yes	0.053
Plant Height														
sc00268ln345453_135128_C_T_166424412	2	38.49	4.39	G	36	59	A	172	55	4	16	17.31		0.062
sc00090ln635406_285885_T_C_84791703	5	40.25	3.04	G	175	55	A	33	60	4	16.1	15.87		0.609
sc00296ln326650_106196_C_A_175840672	8	52.73	4.83	C	173	55	A	35	60	5	20.7	16.83	Yes	0.045
sc00211ln404231_317008_A_G_145539365	n/a	n/a	3.07	G	60	59	A	148	55	4	17.9	28.85	Yes	0.609

Significant marker table (continued).

Markers (Durango)	Ch	Mbp	Log10P	A1	N	Mean	A2	N	Mean	Diff	R2 (%)	MAF	Stepwise	FDR_p
Yield														
sc00013ln1423374_1157977_A_C_22673519	7	42.38	3.38	A	45	2652	C	81	2397	255	8.72	35.7	Yes	0.9088
Maturity														
sc00382ln273856_226597_C_T_201742336	8	11.92	3.04	A	116	106	G	10	111	5	21.4	7.94	Yes	0.5906
sc00186ln436341_356682_A_G_135072564	10	43.04	3.36	A	53	109	G	73	105	4	46.5	42.1	Yes	0.5906
sc00725ln159840_35413_C_T_273109418	11	49.62	3.02	A	10	104	G	116	106	2	5.82	7.94		0.5906
Plant Height														
sc00384ln271106_234031_C_T_202297427	8	54.69	3.29	A	103	56	G	23	58	2	4.02	18.3	Yes	0.4277
sc00384ln271106_266158_T_C_202329554	8	54.72	3.04	A	101	56	G	25	58	2	2.76	19.8		0.4277
sc01132ln102687_11305_T_C_325428675	8	54.76	3.04	A	101	56	G	25	58	2	2.76	19.8		0.4277
sc01782ln58201_36433_A_G_375870656	8	54.89	3.04	A	101	56	G	25	58	2	2.76	19.8		0.4277
Seed Weight														
sc00563ln203201_111347_G_A_244031074	2	37.16	3.18	A	74	36.2	G	52	36.4	0	0.02	41.3		0.3692
sc00038ln842375_617118_G_A_47869458	2	46.34	3.11	A	51	35.0	G	75	37.2	2	6.35	40.5	Yes	0.3692
sc00038ln842375_668515_C_T_47920855	2	46.37	3.11	A	51	35.0	G	75	37.2	2	6.35	40.5		0.3692
sc00038ln842375_820428_T_C_48072768	2	46.52	3.11	A	75	37.2	G	51	35.0	2	6.35	40.5		0.3692
sc00335ln300518_284390_G_A_188281336	3	6.48	4.04	A	14	35.3	G	112	36.4	1	0.69	11.1		0.2592
sc00236ln366267_289928_C_T_155168008	3	9.25	3.02	A	67	37.0	G	59	35.5	2	3.21	46.8		0.3926
sc00324ln307318_62459_C_T_184705118	9	13.31	3.7	A	116	36.6	G	10	32.0	5	8.6	7.94	Yes	0.2837

Significant marker table (continued).

Markers (Mesoamerican)	Ch	Mbp	Log10P	A1	N	Mean	A2	N	Mean	Diff	R2 (%)	MAF	Step wise	FDR_p
Plant Height														
sc01058ln111657_20684_G_A_317550805	5	4.88	3.37	A	76	55	G	6	57	2	14.6	7.32	Yes	0.65
Yield										0				
sc00174ln464616_175582_G_A_129497046	1	0.29	3.32	A	20	95	G	62	88	7	9.41	24.39	Yes	0.021
sc00240ln364462_178258_A_C_156519318	2	3.57	4.72	A	50	89	C	32	92	3	0.18	39.02		0.004
sc00038ln842375_538437_G_A_47790777	2	46.26	4.54	A	7	86	G	75	90	4	4.4	8.54		0.004
sc00038ln842375_519682_G_T_47772022	2	46.24	4.54	A	75	90	C	7	86	4	4.4	8.54		0.004
sc00160ln486724_338720_T_C_122981651	2	44.64	4.46	A	74	90	G	8	88	2	3.52	9.76		0.004
sc00025ln963649_327459_T_C_35785661	2	45.12	4.46	A	8	88	G	74	90	2	3.52	9.76		0.004
sc00160ln486724_301841_G_T_122944772	2	44.61	4.46	A	74	90	C	8	88	2	3.52	9.76		0.004
sc00160ln486724_325087_T_C_122968018	2	44.63	4.46	A	8	88	G	74	90	2	3.52	9.76		0.004
sc00240ln364462_149197_T_C_156490257	2	3.6	4.24	A	31	92	G	51	89	4	0.6	37.8		0.006
sc00113ln562714_437945_A_G_98726289	2	46.67	4.2	A	77	90	G	5	84	7	5.45	6.1		0.006
sc00183ln440474_149829_C_T_133550778	2	26.24	3.73	A	55	89	G	27	92	4	0.35	32.93		0.013
sc00240ln364462_172587_A_G_156513647	2	3.58	3.66	A	32	93	G	50	88	5	0.96	39.02		0.014
sc00137ln512899_439414_C_T_111611698	2	25.85	3.6	A	25	93	G	57	89	4	0.39	30.49		0.015
sc00137ln512899_106457_A_C_111278741	2	25.52	3.6	A	25	93	C	57	89	4	0.39	30.49		0.015
sc00315ln315270_304275_C_A_182148417	2	41.39	3.34	A	76	90	C	6	87	4	3	7.32		0.021
sc00301ln323982_220390_T_C_177581102	2	35.04	3.34	A	6	87	G	76	90	4	3	7.32		0.021
sc01622ln65866_17143_G_A_365960311	2	5.76	3.33	A	44	87	G	38	93	6	2.91	46.34		0.021
sc01016ln116177_107973_T_G_312865638	2	6.66	3.3	A	11	88	C	71	90	3	0.74	13.41	Yes	0.021
sc00137ln512899_418822_T_G_111591106	2	25.83	3.26	A	56	89	C	26	93	4	0.76	31.71		0.021
sc00485ln227615_200282_T_C_227438285	2	25.25	3.26	A	56	89	G	26	93	4	0.76	31.71		0.021
sc00137ln512899_218410_G_A_111390694	2	25.63	3.26	A	26	93	G	56	89	4	0.76	31.71		0.021
sc00137ln512899_266557_A_G_111438841	2	25.68	3.26	A	26	93	G	56	89	4	0.76	31.71		0.021
sc00137ln512899_287286_T_G_111459570	2	25.7	3.16	A	28	93	G	54	89	4	0.65	34.15		0.021
sc00137ln512899_401056_A_C_111573340	2	25.81	3.16	A	28	93	C	54	89	4	0.65	34.15		0.021
sc00102ln598725_310499_T_C_92179296	2	26.85	3.16	A	5	83	G	77	90	7	3.35	6.1		0.021
sc00118ln552338_522535_C_A_101600791	2	28.17	3.16	A	77	90	C	5	83	7	3.35	6.1		0.021
sc00116ln556045_550252_C_T_100518331	2	28.04	3.16	A	5	83	G	77	90	7	3.35	6.1		0.021
sc07321ln2788_2298_T_C_463763859	2	28.05	3.16	A	77	90	G	5	83	7	3.35	6.1		0.021
sc00815ln144488_127459_G_T_286860988	2	28.73	3.16	A	77	90	C	5	83	7	3.35	6.1		0.021
sc00116ln556045_416299_A_G_100384378	2	27.89	3.16	A	5	83	G	77	90	7	3.35	6.1		0.021
sc00129ln528071_25869_C_A_107024247	2	13.63	3.16	A	21	82	C	61	93	11	17.2	25.61		0.021
sc00137ln512899_112454_C_T_111284738	2	25.53	3.03	A	26	92	G	56	89	3	0.35	31.71		0.025
sc00137ln512899_172990_C_T_111345274	2	25.59	3.03	A	26	92	G	56	89	3	0.35	31.71		0.025
sc00137ln512899_252077_T_C_111424361	2	25.66	3.03	A	26	92	G	56	89	3	0.35	31.71		0.025

Significant marker table (continued).

Markers (Mesoamerican)	Ch	Mbp	Log10P	A1	N	Mean	A2	N	Mean	Diff	R2 (%)	MAF	Step wise	FDR_p
Yield										0				
sc00079ln659676_90976_A_G_77428539	3	39.66	3.14	A	46	87	G	36	94	8	11.1	43.9		0.022
sc00779ln149779_105659_A_G_281541703	4	43.58	3.16	A	51	84	G	31	100	16	21.9	37.8	Yes	0.021
sc03228ln22219_1889_C_T_429548590	5	9.05	3.03	A	54	94	G	28	83	11	21.7	34.15		0.025
sc00676ln169706_106063_G_A_265106866	5	4.71	3.01	A	55	93	G	27	83	10	18.7	32.93		0.026
sc02498ln35994_11115_G_A_408566925	7	15.9	7.34	A	24	104	G	58	84	19	42.7	29.27	Yes	0
sc00439ln246418_21317_T_C_216299710	7	14.46	7.34	A	58	84	G	24	104	19	42.7	29.27		0
sc02331ln39589_2360_G_T_402247579	7	14.75	7.34	A	24	104	C	58	84	19	42.7	29.27		0
sc01291ln88348_17155_C_T_340613748	7	15.78	7.34	A	24	104	G	58	84	19	42.7	29.27		0
sc01865ln54877_32766_T_C_380556459	7	35.36	6.56	A	60	85	G	22	104	19	38.9	26.83		1.00E-04
sc00002ln2152649_1976120_C_A_4148171	7	51.59	5.17	A	32	80	C	50	97	17	34.4	39.02		0.002
sc00093ln620690_53709_G_A_86440192	7	45.01	4.55	A	29	97	G	53	86	11	19.1	35.37		0.004
sc00260ln351626_25579_G_A_163521775	7	9.45	4.12	A	59	86	G	23	101	16	30.4	28.05		0.007
sc00260ln351626_31611_G_A_163527807	7	9.45	4.12	A	59	86	G	23	101	16	30.4	28.05		0.007
sc00853ln138233_81612_G_A_292173871	7	39.3	3.82	A	29	78	G	53	97	19	32.4	35.37	Yes	0.012
sc00021ln1025434_384007_C_A_31860445	7	47.58	3.79	A	58	85	C	24	103	18	26.3	29.27		0.012
sc00339ln298666_59263_T_C_189255570	7	47.12	3.79	A	58	85	G	24	103	18	26.3	29.27		0.012
sc00021ln1025434_517061_G_A_31993499	7	47.71	3.69	A	61	86	G	21	102	17	22.7	25.61		0.014
sc00021ln1025434_280672_C_A_31757110	7	47.48	3.43	A	59	85	C	23	103	18	24.3	28.05		0.021
sc00093ln620690_104379_A_G_86490862	7	44.95	3.22	A	31	96	G	51	87	9	15.7	37.8		0.021
sc00067ln694271_455158_G_A_69620416	7	2.66	3.12	A	67	87	G	15	102	14	16.2	18.29		0.022
sc00327ln306820_257821_C_T_185821957	7	10.82	3.07	A	59	86	G	23	100	14	26.6	28.05		0.024
sc00071ln681296_590679_G_A_72513690	8	56.86	3.4	A	55	92	G	27	87	5	4.63	32.93		0.021
sc00071ln681296_656721_T_C_72579732	8	56.92	3.4	A	55	92	G	27	87	5	4.63	32.93		0.021
sc00068ln689895_606631_T_C_70466160	8	57.03	3.39	A	24	92	G	58	89	3	8.22	29.27		0.021
sc00015ln1350335_1224361_T_G_25560637	9	36.71	3.79	A	48	83	C	34	100	18	33.7	41.46	Yes	0.012
sc00015ln1350335_1191524_C_T_25527800	9	36.68	3.43	A	35	100	G	47	83	17	31	42.68		0.021
sc00805ln146292_74900_C_T_285352806	9	19	3.13	A	34	96	G	48	86	10	25.4	41.46		0.022
sc00015ln1350335_1210315_C_T_25546591	9	36.7	3.11	A	33	100	G	49	83	17	31.2	40.24		0.022
sc00119ln552178_203683_C_T_101834277	10	38.18	4.88	A	63	93	G	19	80	14	30.4	23.17	Yes	0.004
sc00309ln320131_212493_G_A_180150638	10	40.62	4.59	A	41	86	G	41	94	8	20.1	50		0.004
sc01642ln64808_3042_T_C_367252976	10	32.65	4.21	A	18	81	G	64	93	12	22.3	21.95		0.006
sc01960ln51199_20704_C_T_385572663	10	32.13	4.21	A	18	81	G	64	93	12	22.3	21.95		0.006
sc01321ln86834_73124_C_A_343299461	10	32.73	4.21	A	18	81	C	64	93	12	22.3	21.95		0.006
sc01573ln68558_33641_T_C_362684843	10	17.22	3.34	A	60	93	G	22	82	11	19.3	26.83		0.021
sc01473ln75007_52419_A_G_355556346	10	17.31	3.3	A	22	82	G	60	93	11	18.8	26.83		0.021
sc01698ln61778_39107_G_A_370834562	10	12.63	3.2	A	62	93	G	20	82	10	16.9	24.39		0.021

Significant marker table (continued).

Markers (Mesoamerican)	Ch	Mbp	Log10P	A1	N	Mean	A2	N	Mean	Diff	R2 (%)	MAF	Step wise	FDR_p
Yield										0				
sc00932ln126303_99263_A_G_302660300	10	18.46	3.16	A	61	93	G	21	82	11	17.2	25.61		0.021
sc01227ln94195_59104_C_T_334822650	10	18.06	3.16	A	21	82	G	61	93	11	17.2	25.61		0.021
sc02111ln45786_24422_C_T_392904062	10	18.39	3.16	A	21	82	G	61	93	11	17.2	25.61		0.021

Significant marker table (continued).

Markers (Mesoamerican)	Ch	Mbp	Log10P	A1	N	Mean	A2	N	Mean	Diff	R2 (%)	MAF	Step wise	FDR_p
Seed Weight														
sc00327ln306820_257821_C_T_185821957	7	10.8	3.65	A	59	20.6	G	23	22.1	1.5	15.73	28.05	Yes	0.1979
sc00327ln306820_246895_C_T_185811031	7	10.8	3.11	A	60	20.7	G	22	22	1.4	13.23	26.83		0.1979
sc00555ln205282_74219_C_T_242362884	7	10.7	3.11	A	60	20.7	G	22	22	1.4	13.23	26.83		0.1979
sc00071ln681296_466314_A_C_72389325	8	56.7	3.49	A	68	20.7	C	14	22.4	1.7	15.03	17.07	Yes	0.1979
sc00805ln146292_74900_C_T_285352806	9	19	3.01	A	34	21.7	G	48	20.5	1.2	12.79	41.46		0.1979
Maturity														
sc00893ln132397_78914_G_T_297593897	1	3.86	4.49	A	39	80	C	43	99	18	48	47.56	Yes	0.003
sc00174ln464616_175582_G_A_129497046	1	0.29	3.37	A	20	95	G	62	88	7	4.72	24.39		0.0164
sc00010ln1529725_1084965_A_G_18124555	1	4.39	3.1	A	43	82	G	39	99	17	39.25	47.56		0.0241
sc00240ln364462_149197_T_C_156490257	2	3.6	5.57	A	31	92	G	51	89	4	2.06	37.8		0.0005
sc00240ln364462_178258_A_C_156519318	2	3.57	5.56	A	50	89	C	32	92	3	1.55	39.02		0.0005
sc00240ln364462_172587_A_G_156513647	2	3.58	4.82	A	32	93	G	50	88	5	2.92	39.02		0.0017
sc00183ln440474_149829_C_T_133550778	2	26.2	4.34	A	55	89	G	27	92	4	1.65	32.93	Yes	0.0036
sc01622ln65866_17143_G_A_365960311	2	5.76	4.19	A	44	87	G	38	93	6	5.19	46.34		0.0047
sc00137ln512899_112454_C_T_111284738	2	25.5	4.12	A	26	92	G	56	89	3	0.99	31.71		0.0047
sc00137ln512899_172990_C_T_111345274	2	25.6	4.12	A	26	92	G	56	89	3	0.99	31.71		0.0047
sc00137ln512899_252077_T_C_111424361	2	25.7	4.12	A	26	92	G	56	89	3	0.99	31.71		0.0047
sc00137ln512899_439414_C_T_111611698	2	25.9	4.04	A	25	93	G	57	89	4	1.59	30.49		0.0054
sc00137ln512899_106457_A_C_111278741	2	25.5	4.04	A	25	93	C	57	89	4	1.59	30.49		0.0054
sc00137ln512899_418822_T_G_111591106	2	25.8	3.76	A	56	89	C	26	93	4	2.39	31.71		0.008
sc00485ln227615_200282_T_C_227438285	2	25.3	3.76	A	56	89	G	26	93	4	2.39	31.71		0.008
sc00137ln512899_218410_G_A_111390694	2	25.6	3.76	A	26	93	G	56	89	4	2.39	31.71		0.008
sc00137ln512899_266557_A_G_111438841	2	25.7	3.76	A	26	93	G	56	89	4	2.39	31.71		0.008
sc00203ln411639_380132_A_G_142338465	2	38.7	3.74	A	73	88	G	9	106	18	17.29	10.98		0.0083
sc00137ln512899_287286_T_G_111459570	2	25.7	3.67	A	28	93	C	54	89	4	2.44	34.15		0.0091
sc00137ln512899_401056_A_C_111573340	2	25.8	3.67	A	28	93	C	54	89	4	2.44	34.15		0.0091
sc00240ln364462_162645_G_A_156503705	2	3.59	3.62	A	31	93	G	51	88	5	3.79	37.8		0.0099
sc00183ln440474_259280_T_G_133660229	2	26.4	3.14	A	26	94	C	56	88	6	4.32	31.71		0.0231
sc00137ln512899_22309_A_G_111194593	2	25.4	3.13	A	55	88	G	27	94	5	3.32	32.93		0.0231

Significant marker table (continued).

Markers (Mesoamerican)	Ch	Mbp	Log10P	A1	N	Mean	A2	N	Mean	Diff	R2 (%)	MAF	Step wise	FDR_p
Maturity														
sc00137ln512899_226303_G_A_111398587	2	25.6	3.13	A	55	88	G	27	94	5	3.32	32.93		0.0231
sc00137ln512899_206565_C_T_111378849	2	25.6	3.13	A	27	94	G	55	88	5	3.32	32.93		0.0231
sc00137ln512899_370382_A_G_111542666	2	25.8	3.13	A	55	88	G	27	94	5	3.32	32.93		0.0231
sc00137ln512899_430228_T_C_111602512	2	25.8	3.13	A	55	88	G	27	94	5	3.32	32.93		0.0231
sc00360ln288358_203794_T_C_195547394	2	25.1	3.13	A	27	94	G	55	88	5	3.32	32.93		0.0231
sc00059ln720534_580545_A_G_64084099	3	1.07	4.32	A	40	98	G	42	82	16	36.17	48.78		0.0036
sc00026ln958753_954186_G_A_37376037	3	46.2	4.17	A	15	93	G	67	87	6	22.82	18.29		0.0047
sc00059ln720534_461954_T_C_63965508	3	1.18	3.42	A	39	83	G	43	97	14	28.77	47.56		0.0148
sc00079ln659676_90976_A_G_77428539	3	39.7	3.11	A	46	87	G	36	94	8	7.92	43.9		0.0239
sc00377ln276023_97213_T_G_200237103	3	0.48	3.09	A	47	89	C	35	92	3	1.24	42.68		0.0241
sc00779ln149779_105659_A_G_281541703	4	43.6	4.4	A	51	84	G	31	100	16	32.53	37.8	Yes	0.0032
sc00853ln138233_81612_G_A_292173871	7	39.3	7.2	A	29	78	G	53	97	19	44.11	35.37	Yes	0
sc02498ln35994_11115_G_A_408566925	7	15.9	6.92	A	24	104	G	58	84	19	44.28	29.27	Yes	0
sc00439ln246418_21317_T_C_216299710	7	14.5	6.92	A	58	84	G	24	104	19	44.28	29.27		0
sc02331ln39589_2360_G_T_402247579	7	14.8	6.92	A	24	104	C	58	84	19	44.28	29.27		0
sc01291ln88348_17155_C_T_340613748	7	15.8	6.92	A	24	104	G	58	84	19	44.28	29.27		0
sc01865ln54877_32766_T_C_380556459	7	35.4	6.55	A	60	85	G	20	104	19	39.4	26.83		0.0001
sc00002ln2152649_1976120_C_A_4148171	7	51.6	6.32	A	32	80	C	50	97	17	37.05	39.02		0.0001
sc00021ln1025434_384007_C_A_31860445	7	47.6	5.71	A	58	85	C	24	103	18	37.51	29.27		0.0004
sc00339ln298666_59263_T_C_189255570	7	47.1	5.71	A	58	85	G	24	103	18	37.51	29.27		0.0004
sc00021ln1025434_280672_C_A_31757110	7	47.5	5.44	A	59	14	C	23	103	89	34.94	28.05		0.0006
sc00262ln351368_226896_C_T_164426117	7	44	5.35	A	39	100	G	43	81	19	50.35	47.56	Yes	0.0007
sc01434ln78222_42548_A_C_352565471	7	43.8	5	A	33	79	C	49	97	18	43.76	40.24		0.0014
sc00021ln1025434_517061_G_A_31993499	7	47.7	4.92	A	61	86	G	21	102	17	29.85	25.61		0.0016
sc00093ln620690_53709_G_A_86440192	7	45	4.85	A	29	97	G	53	86	11	14.99	35.37		0.0017
sc00339ln298666_189468_G_A_189385775	7	47	4.53	A	61	86	G	21	103	17	30.23	25.61		0.003
sc00021ln1025434_407784_C_T_31884222	7	47.6	4.41	A	61	86	G	21	102	17	29.85	25.61		0.0032
sc00021ln1025434_525563_C_A_32002001	7	47.7	3.99	A	63	86	C	19	102	16	25.1	23.17		0.0059
sc00021ln1025434_470964_G_A_31947402	7	47.7	3.96	A	62	86	G	20	102	16	27.25	24.39		0.006
sc00469ln235093_87464_C_T_223609029	7	46.7	3.93	A	20	102	G	62	86	16	27.4	24.39		0.0063
sc00021ln1025434_93724_A_G_31570162	7	47.3	3.77	A	58	85	G	24	102	16	31.6	29.27		0.008
sc00394ln266395_229160_G_A_204982812	7	44.2	3.5	A	47	96	G	35	82	15	29.97	42.68		0.0129
sc00093ln620690_104379_A_G_86490862	7	45	3.42	A	31	96	G	51	87	9	11.6	37.8		0.0148
sc00394ln266395_95537_T_C_204849189	7	44.3	3.29	A	48	83	G	34	100	17	39.29	41.46		0.019
sc00706ln163930_38021_T_G_270038084	7	40.2	3.22	A	26	79	C	56	95	17	34.62	31.71		0.0215
sc03952ln14595_12945_G_T_442698187	7	32.6	3.14	A	48	97	C	34	80	17	40.17	41.46		0.0231

Significant marker table (continued).

Markers (Mesoamerican)	Ch	Mbp	Log10P	A1	N	Mean	A2	N	Mean	Diff	R2 (%)	MAF	Step wise	FDR_p
Maturity														
sc00071ln681296_590679_G_A_72513690	8	56.9	4.5	A	55	92	G	27	87	5	3.43	32.93		0.003
sc00071ln681296_656721_T_C_72579732	8	56.9	4.5	A	55	92	G	27	87	5	3.43	32.93		0.003
sc00015ln1350335_1224361_T_G_25560637	9	36.7	4.47	A	48	83	C	34	100	18	42.94	41.46	Yes	0.003
sc00015ln1350335_1191524_C_T_25527800	9	36.7	3.86	A	35	100	G	47	83	17	39.51	42.68		0.0073
sc00015ln1350335_1210315_C_T_25546591	9	36.7	3.7	A	33	100	G	49	83	17	39.74	40.24		0.0089
sc00309ln320131_212493_G_A_180150638	10	40.6	3.35	A	41	86	G	41	94	8	9.96	50		0.0169
sc00631ln181891_1253_C_T_257082952	Scaff_42	0.02	3.01	A	14	103	G	68	87	16	19.7	17.07		0.029

Significant marker table (continued).

Markers (AYT)	Ch	Mbp	Log10P	A1	N	Mean	A2	N	Mean	Diff	R^2 (%)	MAF	Step wise	FDR_p
Yield														
sc00675ln170111_135617_A_G_264966309	2	42.07	3.83	A	107	2570	C	29	2327	242.77	10.43	21.32	Yes	0.2189
sc00039ln840544_527720_C_T_48622435	3	36.1	3.52	A	37	2463	C	99	2539	75.92	3.8	27.21		0.2968
sc00853ln138233_81612_G_A_292173871	7	39.3	3.04	A	66	2397	G	70	2632	235.06	0.25	48.53	Yes	0.4895
sc00002ln2152649_1976120_C_A_4148171	7	51.59	4.29	A	57	2450	G	79	2568	118.52	6.07	41.91		0.1517
sc01477ln74590_51381_G_A_355854774	9	0.3	3.31	A	37	2420	G	99	2555	135.17	2.32	27.21		0.3587
sc00030ln901868_598673_A_C_40799116	9	25.39	3	A	59	2438	G	77	2580	141.61	3.19	43.38		0.4895
Plant Height														
sc00268ln345453_135128_C_T_166424412	2	38.49	4.22	A	11	53	G	71	55	1.7	1.14	13.97	Yes	0.0356
sc00211ln404231_254799_G_A_145477156	5	34.66	5.49	A	10	53	G	72	55	1.93	0.48	13.24		0.0032
sc00211ln404231_261454_C_T_145483811	5	34.66	5.49	A	72	55	C	10	56	1.22	10.3	13.24		0.0032
sc00090ln635406_285885_T_C_84791703	5	40.25	4.86	A	8	54	C	74	55	0.56	0.95	11.03		0.0103
sc00296ln326650_106196_C_A_175840672	8	52.73	3.08	A	76	55	G	6	56	0.58	0.26	8.09		0.2889
sc00797ln146967_28939_G_A_284133954	8	22.7	3.05	A	73	55	G	9	55	0.21	3.34	11.03	Yes	0.2889
sc08776ln2265_1282_A_C_467399132	11	39.88	3.23	A	12	54	G	70	55	0.77	10.87	15.44	Yes	0.2889
sc00211ln404231_317008_A_G_145539365	n/a	n/a	5.61	A	10	55	G	72	55	0.22	0.96	13.97		0.0032
Maturity														
sc00002ln2152649_1976120_C_A_4148171	7	51.59	5.39	A	57	99	G	79	95	3.78	7.43	41.91		0.012
sc00853ln138233_81612_G_A_292173871	7	39.3	4.64	A	66	97	G	70	96	1.74	1.99	48.53		0.0339
sc00789ln147969_135375_A_G_283059614	8	4.79	3.14	A	111	95	G	25	102	6.58	3.79	18.38	Yes	0.3604
sc00862ln137460_111453_C_T_293443775	8	4.92	3.14	A	111	95	G	25	102	6.58	0.78	18.38		0.3604
sc00862ln137460_62873_C_T_293395195	8	4.87	3.14	A	111	96	G	25	99	2.98	2.69	18.38		0.3604
sc00055ln737569_432814_C_T_61008497	9	29.79	4.43	A	78	101	G	58	90	10.75	8.13	42.65	Yes	0.0365
sc00015ln1350335_1191524_C_T_25527800	9	36.68	3.05	A	80	99	G	56	93	5.23	1.88	41.18	Yes	0.3807
Seed Weight														
sc00293ln329559_110683_C_T_174860214	2	13.25	3.34	A	100	25.0	G	36	32.9	7.83	18.29	26.47	Yes	0.4536
sc00060ln715437_619181_T_G_64843269	6	25.17	3.01	A	70	25.9	C	66	28.4	2.53	2.46	48.53		0.5916
sc00015ln1350335_515184_A_G_24851460	9	36.01	3.51	A	68	33.3	G	68	20.9	12.41	59.01	50	Yes	0.4536
sc00055ln737569_432814_C_T_61008497	9	29.79	3.46	A	58	21.1	G	78	31.6	10.5	41.39	42.65	Yes	0.4536