

GENETIC DIVERSITY OF THE GUATEMALAN CLIMBING BEAN COLLECTIONS

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

By

María Gabriela Tobar Piñón

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Department:
Plant Sciences

December 2017

Fargo, North Dakota

North Dakota State University
Graduate School

Title

GENETIC DIVERSITY OF THE GUATEMALAN CLIMBING BEAN
COLLECTIONS

By

María Gabriela Tobar Piñón

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:

Phillip McClean

Chair

Juan M. Osorno

Kevin McPhee

Jill Hamilton

Approved:

12/12/2017

Date

Richard D. Horsley

Department Chair

ABSTRACT

Since common bean is the most important legume crop for human consumption around the world, bean breeders are challenged to increase production of beans while facing new problems like climate change. Guatemalan climbing beans have been suggested to represent race Guatemala, a newly identified race in the Middle American gene pool that may represent an untapped source of alleles for bean improvement. This study confirmed the existence of race Guatemala in the Middle American gene pool and its differentiation from other races. The low population structure found within these Guatemalan beans also makes this population ideal for discovery of candidate genes for important traits. We demonstrate that the Guatemalan population was useful to provide candidate genes for previously reported genetic factors like the *V* gene for flower color, and the *Asp* gene for seed coat luster. The important relationship between flowering time and altitudinal adaptation of beans was also emphasized.

ACKNOWLEDGEMENTS

I would first like to thank Dr. Phil McClean for his support and guidance during this process. He allowed that this project was my own work but guided me in the right direction whenever I needed it. I would also like to express my gratitude to my committee members Dr. Juan Osorno, Dr. Jill Hamilton and Dr. Kevin McPhee who were always willing to help. I am indebted to them for their invaluable contribution.

I would also like to acknowledge Dr. Samira Mafi Moghaddam, Rian Lee and Dr. Atena Oladzad from the Dry Bean Genomics Lab at NDSU for their instruction and help on each step of this project. Without their guidance, this project would have never been possible to conclude. I must express my gratitude to my colleagues at the Institute of Agricultural Science and Technology (ICTA) in Guatemala, Julio Villatoro, Angela Miranda and Carlos Maldonado for their support and their key role in the project.

I will not forget the support and kindness of USAID, The Legume Innovation Lab Project, NDSU and ICTA, who made possible this great opportunity in my career. Without their help this research would not have been possible.

I will also like to thank to my family for their wise counsel and encourage to pursue my dreams. Finally, thanks to my good friends who supported me in this journey.

DEDICATION

To the memory of Jaime Piñón Gramajo, Estela Bonilla González and Manuel Tobar Arévalo;
and to Elvira Oseida Carcuz.

TABLE OF CONTENTS

| | |
|--|-----|
| ABSTRACT..... | iii |
| ACKNOWLEDGEMENTS..... | iv |
| DEDICATION..... | v |
| LIST OF TABLES..... | ix |
| LIST OF FIGURES..... | x |
| LIST OF APPENDIX TABLES..... | xi |
| LIST OF APPENDIX FIGURES..... | xii |
| INTRODUCTION..... | 1 |
| LITERATURE REVIEW..... | 5 |
| Common Dry Bean..... | 5 |
| Nutritional content..... | 5 |
| Climbing beans..... | 6 |
| The <i>milpa</i> system..... | 6 |
| Origin and Domestication of Common Bean..... | 7 |
| Diversity of Common Bean..... | 10 |
| Common Bean in the U.S..... | 13 |
| Common Bean in Guatemala..... | 15 |
| The Genome of Common Bean..... | 16 |
| Molecular markers for diversity analysis..... | 17 |
| Genotyping-by-sequencing (GBS)..... | 18 |
| SNP calling..... | 19 |
| Association mapping..... | 19 |
| MATERIALS AND METHODS..... | 21 |
| Genetic Diversity and Population Structure Analysis..... | 21 |

| | |
|---|-----------|
| Plant material..... | 21 |
| Tissue collection and DNA extraction | 23 |
| Genotyping-by-sequencing (GBS) | 23 |
| Genotypes separation and barcodes trimming..... | 24 |
| Quality control..... | 24 |
| Mapping and read group information..... | 25 |
| SNP calling..... | 25 |
| Data imputation | 25 |
| Population structure analysis..... | 26 |
| Genetic diversity analysis..... | 27 |
| Geographical distribution in the GUA_2015 collection..... | 28 |
| Intra-accession diversity analysis | 28 |
| Genome-Wide Association Study (GWAS)..... | 30 |
| RESULTS | 33 |
| Genetic Diversity and Population Differentiation..... | 33 |
| SNP calling..... | 33 |
| Genetic diversity analysis..... | 33 |
| Intra-accession diversity analysis | 34 |
| Population differentiation..... | 35 |
| Population structure analysis..... | 39 |
| Geographical structure of the genetic variation in population GUA_2015..... | 44 |
| Genome Wide Association Study | 45 |
| Genomics regions associated with altitudinal adaptation in population GUA_2015..... | 47 |
| Genomics regions associated with seed length, seed width and seed width/length ratio in population GUA_1966-82..... | 48 |
| Genomic regions associated with flower color, cotyledon color and stem color..... | 49 |

| | |
|--|----|
| Genomic regions associated with seed coat luster in population GUA_1966-82 | 51 |
| Genomic regions associated with pod color in population GUA_1966-82..... | 52 |
| Genomic regions associated with rust (<i>Uromyces appendiculatus</i> (Pers) Unger) resistance in population GUA_1966-82 | 53 |
| Genomic regions associated with anthracnose (<i>Colletotrichum lindemuthianum</i>) resistance in population GUA_1966-82 | 54 |
| Genomic regions associated with downy mildew (<i>Phytophthora nicotianae</i>) resistance in population GUA_1966-82 | 55 |
| DISCUSSION | 57 |
| Genetic Diversity and Population Structure | 57 |
| Genome Wide Association Study | 61 |
| Altitudinal adaptation | 61 |
| Seed shape | 65 |
| Flower color, cotyledon color and stem color | 68 |
| Seed coat luster | 70 |
| Pod color..... | 71 |
| Rust resistance | 72 |
| Anthracnose resistance | 75 |
| Downy mildew resistance..... | 76 |
| CONCLUSIONS..... | 79 |
| REFERENCES | 80 |
| APPENDIX..... | 95 |

LIST OF TABLES

| <u>Table</u> | <u>Page</u> |
|---|-------------|
| 1. Phenotypic data for the GUA_1966-82 population. | 32 |
| 2. Summary of diversity statistics of the different populations. | 34 |
| 3. Analysis of molecular variance (AMOVA) among and within accessions. | 35 |
| 4. Pairwise F_{ST} values of all the populations evaluated..... | 36 |
| 5. AMOVA for separation of gene pools, Middle American races and Guatemalan groups..... | 37 |
| 6. Summary of intra-accession diversity statistics for GUA_1966-82 genotypes..... | 38 |
| 7. Mean Square Deviation (MSD) for each GWAS model and each trait evaluated..... | 46 |

LIST OF FIGURES

| <u>Figure</u> | <u>Page</u> |
|--|-------------|
| 1. Phylogenetic tree for the races published by Blair et al. (2009) with 604 accessions of common bean..... | 13 |
| 2. Map of the Guatemalan western highlands where climbing bean seed from GUA_2015 was collected. | 22 |
| 3. ΔK results showing the best $K=3$ | 40 |
| 4. Results of population structure analysis with K values ranging from 2 to 5. | 41 |
| 5. Results of PCA for the first three principal components based on the SNPs information..... | 42 |
| 6. Results of the maximum likelihood tree. | 43 |
| 7. Results of PCA for the first two principal components. | 45 |
| 8. a) Manhattan plot for altitudinal adaptation of GUA_2015. b) quantile-quantile plot for the best model. | 47 |
| 9. a) Manhattan plots for seed length, seed width and seed width/length ratio in accessions from GUA_1966-82. b) quantile-quantile plots for the best models. | 49 |
| 10. a) Manhattan plots for flower color, cotyledon color and stem color in accessions of GUA_1966-82. b) quantile-quantile plots for the best models. | 51 |
| 11. a) Manhattan plot for seed coat appearance in accessions of GUA_1966-82. b) quantile-quantile plot for the best model. | 52 |
| 12. a) Manhattan plot for pod color in accessions of GUA_1966-82. b) quantile-quantile plot for the best model. | 53 |
| 13. a) Manhattan plot for rust resistance in accessions of GUA_1966-82. b) quantile-quantile plot for the best model. | 54 |
| 14. a) Manhattan plot for anthracnose resistance in accessions of GUA_1966-82. b) quantile-quantile plot for the best model. | 55 |
| 15. a) Manhattan plot for downy mildew resistance in accessions of GUA_1966-82. b) quantile-quantile plot for the best model. | 56 |

LIST OF APPENDIX TABLES

| <u>Table</u> | <u>Page</u> |
|--|-------------|
| A1. InDel markers used in the intra-accession analysis. | 96 |
| A2. Significant markers at 0.01 percentile and candidate genes for each trait evaluated by GWAS. | 99 |

LIST OF APPENDIX FIGURES

| <u>Figure</u> | <u>Page</u> |
|---|-------------|
| A1. Results of the maximum likelihood tree including all the populations analyzed. | 104 |
| A2. Manhattan plots and QQ plots for the best models of the eight agronomic traits that were not analyzed in the discussion..... | 105 |

INTRODUCTION

Common bean (*Phaseolus vulgaris* L.) is the most important pulse crop for human consumption around the world (Broughton et al., 2003), and 46% of the global area for legume production is represented by common bean (Akibode and Maredia, 2012). The production of beans is more than 23 million tons worldwide and almost a third of it is produced by smallholders from developing countries of Africa and Latin America (Broughton et al., 2003). According to Beebe et al. (2000), common beans from the Middle American gene pool constitute more than 60% of the production in the world. In Central America, beans are considered the most profitable crop among the traditional crops especially when compared to cereals. However, income may vary due to the small farm size used for production in this area of the world (Broughton et al., 2003).

Guatemala is a Central America country with the highest chronic malnutrition among all Latin American countries (Marini and Gragnolati, 2003). Malnutrition primarily affects native children 5 years of age or less (Osorno and McClean, 2013). In Guatemala, common bean per capita consumption is 9.4 kg per year (Osorno and McClean, 2013), and it represents 11% of per capita protein intake (Akibode and Maredia, 2012), making it the most important protein source in the country. Dry bean seed contain about 20% protein, and protein content is higher in black beans than kidney and pinto beans (Akibode and Maredia, 2012; Broughton et al., 2003). Common bean production represents 17.8% of all the available crop production area in Guatemala (MAGA, 2013). In the western highlands of the country, climbing beans are intercropped with maize (*Zea mays* L.) and/or squash (*Cucurbita* spp.) in a cropping system known as milpa (Zizumbo-Villareal and Colunga-Garcia, 2012). The primary species of Guatemalan climbing beans is *P. vulgaris*. *P. dumosus* (syn. *P. polyanthus*) and *P. coccineus* are

two other climbing bean species grown in the milpa system. The seeds of *P. vulgaris* are mostly black and round and have been used by small farmers in this region for generations.

Common bean originated in Mexico (Bitocchi et al., 2012) and is distributed from Mexico to Argentina (Gepts et al., 1986). Two independent domestication events gave rise to the Middle American and Andean gene pools (Gepts et al., 1986; Singh et al., 1991a, b).

Historically, each gene pool was divided into three different races: Mesoamerica, Durango, and Jalisco in the Middle American gene pool and; Peru, Chile and Nueva Granada in the Andean gene pool (Singh et al., 1991a). These races represent most of the diversity known in common bean. However, Beebe et al. (2000) suggested that Guatemalan climbing beans are genetically distinct from the six races and should be considered a unique race. Other studies also support the existence of a race Guatemala (Chacón et al., 2005; Blair et al., 2006; Diaz and Blair, 2006; Blair et al., 2009; Blair et al., 2013).

Genetic diversity measures the variation within a population, which has a fundamental role in the evolution of that population (Hughes et al., 2008). To preserve population diversity, it is necessary to understand how genetic diversity is distributed across a species (Loveless and Hamrick, 1984). The current population structure of a collection of individuals within a species reflects evolution over space and time of this population and is important to our understanding of how the species evolved. Evaluating genetic diversity and population structure is necessary to improve a population through plant breeding because it informs decisions such as parental selection and long-term conservation of important germplasm (Acosta-Gallegos et al., 2007; Bitocchi et al., 2012). From this perspective, understanding the genetic structure of Guatemalan climbing beans relative to other races will support decisions regarding this germplasm as a potential source of new alleles for plant breeding.

Common bean diversity analyses have used different types of molecular markers such as random amplified polymorphic DNA (RAPD), single nucleotide polymorphism (SNP) and single sequence repeat (SSR) (Beebe et al., 2000; Ponciano-Samayoa et al., 2009; McClean et al., 2012; Blair et al., 2013; Mercati et al., 2013). SNPs are a recent tool that provides the most abundant set of markers across the genome (Blair et al., 2013). It has been reported that SNPs detect lower polymorphism compared with SSRs, due to the higher number of alleles in SSRs. However, SNPs can detect higher heterozygosity, allow for the generation of denser genetic maps, and are useful for association mapping studies (Blair et al., 2013; Müller et al., 2015). In their study, Müller et al. (2015) concluded that three times more SNPs than SSRs are needed to generate the same diversity information. Currently, it is relatively inexpensive to generate a large number of common bean SNPs using methods such as genotype-by-sequencing (Schröder et al., 2016). The availability of the reference genomes (Schmutz et al., 2014; Vlasova et al., 2016) enables the accurate physical placement of the SNPs in the genome.

Genome-wide association studies (GWAS) associate a phenotype relative to a specific physical position in the genome. It has several advantages relative to bi-parental mapping (Li et al., 2013) principally the ability to evaluate a larger number of recombinant events. GWAS is based on the linkage disequilibrium (LD) of the population; the higher LD, the lower the resolution necessary to detect genomic regions associated with the trait of interest (Rossi et al., 2009). Many LD studies in common bean have concluded that population structure is the most important cause of LD (Rossi et al., 2009; Mamidi et al., 2011a; Mamidi et al., 2013; Müller et al., 2015). For this reason, GWAS should be performed within genotypes of the same gene pool or same race.

Ponciano-Samayoa et al. (2009) characterized a Guatemalan climbing bean collection using six microsatellites with the objectives of identifying duplicated lines and detecting significant genes. The lines were divided into twelve major groups, and 46.8% of the lines were duplicates. They recommended the use of at least 30 markers for a complete characterization of the population. The population was morphologically characterized by Orellana et al. (2006). They used 34 varietal descriptors (12 quantitative and 24 qualitative) from the International Center for Tropical Agriculture (CIAT) and identified groups with 71% similarity.

The primary objectives of this research were: 1) to analyze the genetic diversity, population structure and population differentiation of two Guatemalan climbing bean collections (GUA_1966-82 and GUA_2015) of *P. vulgaris*, using SNP markers; 2) to perform a GWAS to locate genomic regions associated with a large collection of phenotypes; and 3) identify candidate genes associated with altitudinal adaptation, seed shape and size, flower color and disease resistance of the Guatemalan climbing beans. Diversity analysis of this germplasm will allow a confirmation of Guatemalan highland beans as another race within the Middle American gene pool. These results will inform decisions regarding the best management practices for this population as its use by common bean breeding programs in Guatemala and the United States.

LITERATURE REVIEW

Common Dry Bean

Worldwide, common bean (*Phaseolus vulgaris* L.), is the most important pulse crop for human consumption and for many countries in Latin America, it is the main source of protein (Broughton et al., 2003). Common bean originated in Mexico (Bitocchi et al., 2012) and is adapted to a large range of environments (Singh et al., 1991b). It belongs to the Fabaceae family (Kelly, 2010). Common bean has a taproot and many secondary roots, with nodules developed from an association with the nitrogen-fixing bacteria *Rhizobium*. Four types of growth habit have been described in common bean: bushy or determinate upright (I), indeterminate upright (II), semi-climbing (III), and strong climbing (IV). The leaves are trifoliolate. The flowers have a tubular calyx, five unequal petals, ten stamens, and one receptive stigma which could be white, lilac, purple, or bicolor. Fruits are pods, and the seeds have two cotyledons. Inside pods are protein-rich seeds (Clavijo, 1980).

Nutritional content

Within pulses, common bean is considered of high nutritional value. It consists of 20-25% protein, mostly phaseolin, the main storage protein of common bean (Broughton, 2003). Generally, phaseolin has a deficiency in the amino acid methionine. For this reason, a bean diet needs to be complemented with the amino acids provided by a carbohydrate rich grain that contains the methionine amino acid. Akibode and Maredia, (2012) reported that the pinto and kidney market classes of beans have around 22% protein, while black beans had around 24% of protein. Common bean is also an important source of potassium, magnesium and iron (Broughton, 2003). In their study, McClean et al. (2017) found that race Mesoamerica has on average a higher concentration of micro and macro elements compared to race Durango.

Climbing beans

Climbing beans with the type IV growth habit can reach over three meters in height. They are found in the lowlands of Latin America in the case of race Mesoamerica, the humid highlands of Mexico and Guatemala in the case of race Jalisco, and the Andean highlands in the case of race Peru (Singh et al., 1991a). Climbing beans that grow in the western highlands of Guatemala are characterized by having late maturity compared to the climbers from Mexico (180 days) and by having spherical seeds (Beebe et al., 2000). These climbing beans, known as “bolonillos”, can grow at 3200 meters above sea level (masl) and mostly have black seeds, but genotypes with red and white seeds are also found. In this geographical region common bean is also intercropped with maize and other crops (Lepiz-Ildefonso et al., 2009).

The *milpa* system

The milpa agro-ecosystem was developed by natives of Mesoamerica and consists of an array of maize, climbing beans and squashes (Ebel et al., 2017). Maize is used as a support for the growth of the beans, beans fix nitrogen in the soil, and squash covers the soil secreting substances that are poisonous for the insects. The three crops also provide complementary dietary components. Maize is rich in carbohydrates, beans provide proteins, and squashes provide lipids. It is suggested that the complementary nutrition was the driver to develop the system and to domesticate the species around 5500-4300 years before present (Zizumbo-Villareal et al., 2012). The milpa system is still used today by farmers in Mesoamerica (Ebel et al., 2017). Families who live in the western highlands of Guatemala use the system to grow maize and climbing beans, and these crops represent the source of their basic diet.

Origin and Domestication of Common Bean

Common bean diverged from soybean (*Glycine max* L. Merr.) and cowpea (*Vigna unguiculata* L. Walp.) an estimated 18 million years and 8 million years ago, respectively (Kelly, 2010). There was a whole genome duplication within the legume lineage that occurred before the divergence of common bean and soybean, and this duplication event took place ~55 million years ago (Schmutz et al., 2014). The wild ancestral forms of common bean are viny plants found in the highlands of Middle America and the Andes (Gepts and Debouck, 1991). The wild ancestral population diverged around 165,000 years ago to form the Middle American and the Andean wild gene pools (Schmutz et al., 2014).

There are five domesticated species within the *Phaseolus* genus. The species listed in order of genetic similarity to common bean are *P. vulgaris*, *P. coccineus* (scarlet runner bean), *P. dumosus* syn. *P. polyanthus* (year bean), *P. acutifolius* (tepariy bean), and *P. lunatus* (lima bean) (Kelly, 2010). Nine clades were defined within the *Phaseolus* genus. However, the five domesticated species come from only two of the clades (*P. vulgaris* and *P. lunatus*) (Delgado-Salinas et al., 1999). Common, scarlet runner, year, and tepary bean all belong to the *P. vulgaris* clade, and Lima bean belongs to the *P. lunatus* clade.

Many theories about the center of origin of common bean as a species have been published. Kami et al. (1995) suggested the western slope of the Andes in Ecuador and northern Peru was the center of origin of common bean, based on the phaseolin sequences of wild types that had no tandem repeats compared to the predominant phaseolin sequence found with other bean germplasm from other regions. This discovery was made in a wild bean population that is almost extinct and has an intermediate phaseolin pattern between Middle American and Andean gene pools. The authors suggested that beans were then disseminated northwards to areas

including Colombia, Central America, and Mexico to develop the Middle American gene pool, and also southwards to Peru, Bolivia and Argentina to form the Andean gene pool. However, Bitocchi et al. (2012) suggested that the origin of common bean is in Mesoamerica, specifically in the Oaxaca Valley of Mexico, which would mean that the Andean gene pool arose after migration from Mesoamerica. Interestingly, Rendón-Anaya et al. (2017) suggested that the Ecuador and Northern Peru wild population described by Kami et al. (1995), diverged from the ancestral *P. vulgaris* population before it underwent the split into the Middle American and Andean gene pools.

Population structure and phylogeny analyses as well as evidence of a bottleneck in the Andes before domestication, supports the theory of a Mesoamerican origin of common bean (Bitocchi et al., 2012). Rossi et al. (2009) supported the occurrence of a substantial bottleneck in the Andean gene pool and explained that Phaseolin type I could become extinct in Mesoamerica. Schmutz et al. (2014) suggested that wild Andean gene pool derived from the wild Middle American population, supporting the Mesoamerican origin as well as the strong bottleneck in the Andean gene pool. The genetic diversity studies of Rodríguez et al. (2016) and Rendón-Anaya et al. (2017) also supported the Oaxaca valley origin.

It has been suggested that the Middle American and Andean gene pools were domesticated at least once around 8,000 years ago in Mexico and South America, respectively (Mamidi et al., 2013). The number of domestication events for both gene pools have been widely discussed (Singh et al., 1991a; Beebe et al., 2000; Papa and Gepts, 2003; Chacón et al., 2005; Kwak and Gepts, 2009; Mamidi et al., 2013; McClean et al., 2012; Schmutz et al., 2014; Rodríguez et al., 2016). Nowadays, the single domestication theory is accepted for each gene pool (Mamidi et al., 2011), and the population bottleneck associated with the appearance of the

Andean gene pool lasted ~76,000 years. The Andean population began expanding ~90,000 years ago, and this growth still continues today. In comparison, there was not a bottleneck observed with the wild Middle American gene pool (Schmutz et al., 2014).

Genetic diversity was significantly reduced during domestication, which is commonly associated with a founder effect (Bitocchi et al., 2013; Mamidi et al., 2013; Schmutz et al., 2014). The reduction in genetic diversity was more pronounced for the Middle American domestication event than for the Andean (Blair et al., 2013). Wild common bean could have adaptive traits that are not present in the domesticated forms due to domestication and the associated bottleneck (Acosta-Gallegos et al., 2007). Domesticated common beans are very similar to the wild relatives still found in Mexico, Central America, and the Andean region (Porch et al., 2013). The main differences are that the domesticated forms of common bean were selected for larger seeds, non-shattering pods, and an upright bush architecture.

Schmutz et al. (2014) determined that the Middle American wild populations were more diverse than Middle American landrace populations, however, the wild Andean population was less diverse than the Andean landraces. They reported 1,835 domestication-associated candidate genes for the Middle American gene pool, and only 748 Andean domestication-associated candidate genes. Along with common bean domestication came changes in morphological and physiological characteristics. Two major changes associated with domestication were larger seed and plant size (Gepts and Debouck, 1991). There are several Middle American candidate genes involved in common bean domestication that are associated with plant size (Schmutz et al., 2014), and there are also 15 Middle American candidate genes that are known to be involved in seed weight.

Photoperiod sensitivity is another trait that differs between wild and cultivated forms of dry bean. The wild forms typically are photoperiod sensitive. However, photoperiod sensitivity decreased as cultivation continued (Gepts and Debouck, 1991; Broughton et al., 2003). Cultivars of common bean also tend to flower earlier than the wild forms (Gepts and Debouck, 1991). Flowering genes are commonly involved in domestication events (Olsen and Wendel, 2013), and there are 25 Middle American and 13 Andean candidate domestication genes involved in flowering (Schmutz et al., 2014).

Other traits associated with dry bean domestication include growth habit, pod shattering, and permeability of the seed coat (Gepts and Debouck, 1991). In dry bean, the wild form has an extreme intermediate climbing growth habit whereas the domesticated forms can have determinate bush growth habits. Pod shattering, another trait affected by domestication, is an undesirable trait, and the wild types of common bean have highly fibrous pod walls that facilitate the shattering. The pods of domesticated cultivars pods lack the fibers and do not shatter, which is a beneficial trait for harvesting seed. Dormancy and water permeability are also essential for germination. The wild forms of common bean have water impermeable seed coats, and the cultivated forms have a water permeable seed coat which helps with germination.

Diversity of Common Bean

Singh et al. (1991a) performed a detailed study that identified six races of common bean (three for each gene pool) based on associations of molecular (phaseolin and allozymes) and morphological markers, phenotypical traits, reproductive factors, ecological adaptation, and geographical distribution of related landraces of common bean. The Middle American races are found in Mesoamerica and include race Mesoamerica found from Mexico through Central America and in Colombia, Venezuela, and Brazil. Race Durango is located in the semiarid

highlands of Mexico and the southwestern US, and race Jalisco is found in highlands of Mexico and Guatemala. The Andean races are Nueva Granada (located in Los Andes in Colombia, Ecuador and Peru; and in Argentina, Belize, Bolivia, Brazil, Chile, Panama and The Caribbean), Chile (found in the dry Southern Andes), and Peru (found in highlands of Colombia to Argentina).

After this initial determination of gene pools and races, many common bean genetic diversity studies were performed using molecular markers, isozymes, phaseolin and morphological traits (Velasquez and Gepts, 1994; Sonnante et al., 1994; Galvan et al., 2003; Kwak and Gepts, 2009; Bitocchi et al., 2012; Blair et al., 2013). These studies generally confirmed this common bean race structure. However, molecular markers have allowed a more specific description of this diversity.

Beebe et al. (2000) evaluated the structure of the Middle American races using RAPD markers. They found that race Mesoamerica divides into two sub-races, M1 and M2, consistent with differences in plant habit, isozymes and resistance to several diseases. Black beans seem to be related to M1, and red beans seem to be related to M2. Race Jalisco was not divided into sub-populations, and most of the accessions were from the Oaxacan highlands. Race Durango was divided into two sub-races: D1 and D2 found in different regions. Beebe et al. (2000) also proposed a fourth race in the Middle American gene pool, which was named race Guatemala. Accessions of this race were from the highlands of Chiapas (Mexico) and Guatemala. It is believed that a lack of migration led to the creation of this race with distinctive characteristics. This race included most of the Guatemalan climbing beans, with accessions resistant to some races of angular leaf spot (ALS) caused by *Pseudocercospora griseola*. Based on a chloroplast

DNA analysis, all the Guatemalan climbing beans evaluated belong to the Haplotype I. Haplotype I was not found in any other race of common bean (Chacon et al., 2005).

Diaz and Blair, (2006) performed a diversity analysis in the Middle American gene pool genotypes. They questioned the existence of race Guatemala due to the limited number of Guatemalan accessions that were classified. However, when Blair et al. (2013) analyzed the diversity of another population of Middle American and Andean beans with a set of 736 SNPs, they found that the two gene pools are highly differentiated and that race Guatemala was highly differentiated from the Mesoamerican sub-races. They also found the diversity of the Middle American gene pool was larger when compared to the Andean gene pool.

Blair et al. (2009) evaluated the diversity of a core collection of common bean using microsatellites. Their results are in agreement with Beebe et al. (2000). They divided race Mesoamerica into two sub-races: M1 and M2 based on seed color. However, races Durango and Jalisco formed a Durango-Jalisco complex rather than being separate races. The complex also divided into two sub-races: DJ1 and DJ2. A third group represented the Guatemalan climbing beans with black or cream-colored seeds. While the race Guatemala was the most diverse race, the diversity was not associated to a specific phaseolin allele or seed width. Figure 1 shows the neighbor-joining tree developed by Blair et al. (2009).

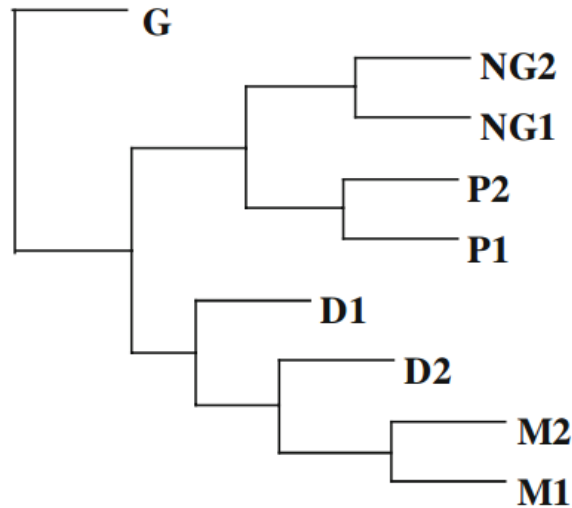


Figure 1. Phylogenetic tree for the races published by Blair et al. (2009) with 604 accessions of common bean.

D= Durango-Jalisco complex (n=108), G= race Guatemala (n=61), NG= race Nueva Granada (n=115), P= race Peru (n=186), M= race Mesoamerica (n=134).

Common Bean in the U.S.

After domestication, common bean was introduced around the world. The currently accepted dissemination pathway is based on phaseolin diversity patterns. A possible dissemination pathway explains that the Middle American and Andean landraces were disseminated to South America and Africa, but Middle American races were predominant in Africa, Europe, and the northeastern portion of the United States (Kelly, 2010). Domesticated beans were introduced to the U.S. by trade. Bean is known as a good source of protein, which is the reason why it is believed that beans expanded into the United States. This expansion was facilitated by native people who lived around the Great Lakes. Market classes of beans typically grown in the U.S. include: pinto (Durango race), which after domestication were presumably moved to central U.S. by native people all the way to the Canadian province of Alberta. Navy and black beans (Mesoamerican race) were moved from the Caribbean to the east coast and then to the Great Lakes region in the US, and kidney beans (Nueva Granada race) were transported

via slave trade by Europeans to Europe and East Africa, then early settlers returned kidney beans to the east coast of the United States (Kelly, 2010).

Common bean cultivation dates back to 2,300 years ago in the southwestern part of the United States (Kaplan, 1965). Dry bean breeding programs in the United States were introduced to improve market classes grown in different regions (Vandermark et al., 2014). Michigan State University started the first dry bean breeding program in the early 1900s (Kelly, 2010). Other universities with dry bean breeding programs include the University of Nebraska, Colorado State University, University of Idaho, and most recently, North Dakota State University. The dry bean breeding program began in the early 1980s at North Dakota State University.

There are also USDA dry bean breeding programs located in Washington, Maryland, Michigan, and Puerto Rico (Kelly, 2010). Different breeding programs in different areas of the United States were known for specializing in certain market classes (Vandemark et al., 2014). For example, navy beans were bred in Michigan, kidney market classes in New York, pinto bean in Colorado and the Pacific Northwest, great northern in Nebraska, small red and pink beans in Washington, and pink bean in California. The preferred market classes in the United States include pinto, black, great northern, red kidney, and navy (Singh, 1999).

Initially, common bean breeding in the U.S. focused on breeding for resistance to *Bean common mosaic virus* (BCMV) and *Beet curly top virus* (BCTV) (Singh et al., 2007). Around the 1970s, when breeding efforts expanded westward, exotic germplasm began to be introgressed to produce cultivars with the upright Type II growth habit. Breeding of common bean tends to be within market classes to maintain the specific characteristics of each market class (Moghaddam et al., 2014). Due to breeding within market classes, the variation within market classes is limited.

Although beans are a good source of protein, only about 14 percent of individuals in the United States eat dry edible beans on any given day (USDA-ERS, 2016). The United States is the second leading producer of dry beans behind only Brazil (Singh, 1999). Currently, North Dakota is the largest producer of commercial dry beans in the United States but that was not always the case (Kelly, 2010). New York was the original state where the dry edible bean industry began in the mid-1800s (USDA-ERS, 2016). Michigan then took the lead in the early 1900s and was the largest dry bean producer until the early 1990s. Since 1994 North Dakota has been the leading producer of beans in the US.

There were 672,588 ha of dry beans produced across the entire United States in 2016 (USDA-ERS, 2016). North Dakota alone accounted for 252,929 of those hectares across 1,682 farms. Most dry bean production in North Dakota takes place around the Red River Valley. Pinto and navy beans are the two main market classes grown in North Dakota accounting for 70% and 17% of production in 2007, respectively. North Dakota also leads the nation in the production of both of these market classes.

Common Bean in Guatemala

Guatemala has the highest rate of chronic malnutrition in Latin America. Maize is the basis of all diets and is the primary source of carbohydrates while bean is the primary source of protein. In Guatemala, common bean is the second most important crop (Aldana, 2010), and black beans are the most consumed market class in the country. The national production in 2014 was 225,760 ton on an area of 243,040 ha. The average seed yield was 928.57 kg·ha⁻¹ in 2013 (MAGA, 2013). The production was distributed within 292,961 producers with an average of 0.82 ha for each producer (IICA, 2014).

The main areas for the production of common bean in Guatemala are the North region in the department of Petén and South East region in the departments of Jutiapa, Chiquimula, and Santa Rosa, where bush type beans are grown. Climbing beans are located mostly in the western highlands of Guatemala, generally in association with maize (Aldana, 2010).

Drought is a problem reported in the production of climbing beans in Guatemala, and in the last few years a 70% drop in yield was reported due to limited rain fall (SESAN, 2014). The most common diseases of common bean in Guatemala are: anthracnose (*Colletotrichum lindemuthianum* (Sacc and Magn.) Scrib), rust (*Uromyces appendiculatus* (Pers: Pers.) Unger), angular leaf spot [*Phaeoisariopsis griseola* (Sacc.)], ascochyta leaf spot (*Phoma exigua* Desmaz.), bacterial brown spot (*Pseudomonas syringae* van Hall.), and some viruses (Schwartz, 2005). Climbing beans are mostly associated with corn, and an aggressive growth is undesirable (Ponciano-Samayoa et al., 2009). Finally, some climbing bean accessions are harvested 180 days after planting. This characteristic is undesirable because it increases the likelihood of damage due to frost during winter.

The Genome of Common Bean

Common bean is a diploid species with eleven chromosomes ($2n=22$). The genome size is around 587 Mb pairs (Schmutz et al., 2014), and 49.2% of the genome is repetitive sequences. The Andean genotype G19833 known as “Chaucha Chuga” and the Mesoamerican genotype BAT93 were used to develop reference genomes for common bean (Schmutz et al., 2014 https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Pvulgaris; Vlasova et al., 2016). The protein coding genes that have been reported in both genome annotations are 27,197 for the genotype G19833 and 30,491 for the genotype BAT93.

Molecular markers for diversity analysis

Molecular markers can identify genotypic differences without analyzing the entire genome, and can be used to identify regions of the genome associated with specific phenotypic traits. They can be defined as DNA fragments that function as “sign posts” and identify information of interest. Also, molecular markers are useful in many biological fields of study. For these reasons, new techniques for their development have emerged (Agarwal et al., 2008). To evaluate diversity, the first DNA-based molecular markers used were the restriction fragment length polymorphism (RFLP). However, the most common markers used in this field are the RAPD, SSR, SNPs, and more recently transposable elements based markers (Poczai et al., 2013). SNPs are the most abundant markers because they represent the smallest unit in a gene that can change. The advances in sequencing have made the use of SNPs as markers possible.

Chen et al. (2013) developed a SSR marker map in common bean with the objective to evaluate genetic diversity and facilitate gene cloning. They found 90 genomic SSR markers and utilized 85 primers into the bean map. They crossed the lines Hong Yundou (known as a standard Chinese landrace with high yield, quality, and resistance to Anthracnose) and Jingdou. In this study, the F₂ segregating population was used to develop the map. The markers were distributed across the eleven chromosomes of common bean. Galeano et al. (2012) developed 313 gene-based SNPs markers, and 53 of these markers were used to develop a molecular map using the recombinant inbred lines (RILs) population of the cross of DOR364 (control genotype) x G19833 (Andean genotype). This map was used to form a consensus map that merged two linkage maps developed from the crosses of BAT93 x JALO EEP558 and DOR364 x BAT477. There were 1060 markers identified in the map, and the length of the map was 2042 cM across

the eleven chromosomes. They concluded these results are useful for population structure analysis and association mapping.

Another type of molecular marker recently developed is based on insertion-deletion events. They are abundant across the genome and less common compared to SNPs. They can also be screened using PCR techniques and have a lower rate of genotyping error compared with the SSRs (Vali et al., 2008). InDels are codominant markers that have been successfully used in diversity studies (Bhatramakki et al., 2002; Vali et al., 2008; Moghaddam et al., 2014). Moghaddam et al. (2014) developed a set of 2687 InDel markers for the market classes of common bean that have demonstrated to be polymorphic and easily amplified.

Genotyping-by-sequencing (GBS)

Considering the interest in discovering gene sequences that control phenotypes in organisms, the development of technologies that make sequencing more efficient has recently been a prime interest. Since the 1990's, Sanger sequencing has helped in genomic advances. However, in the last years, next-generation sequencing (NGS) technologies, such as Illumina's sequencing-by-synthesis (Illumina.com), have allowed the sequencing of whole genomes more efficiently (Mardis, 2008). Now that sequencing is faster and more accurate, the use of sequences to identify single nucleotide polymorphisms is a common tool. However, sequencing plant genomes, which have a high number of repetitive sequences, has been limited by time and cost. Importantly, most genes are located in low copy regions (Elshire et al., 2011). Genotyping-by-sequencing (GBS) is an approach that reduces complexity in a genome, using restriction enzymes that target low copy regions and then sequence just those regions (Poland et al., 2012). GBS was first performed by Elshire et al. (2011) in maize and barley (*Hordeum vulgare*), and is based on the creation of DNA libraries using restriction enzymes and polymerase chain reaction

(PCR) to amplify a subset of DNA fragments. Libraries are then sequenced. Schröder et al. (2016) improved the quality of GBS data for *P. vulgaris*, using the *MseI* and *TaqI* restriction enzymes, which allowed the identification of more SNPs, because the enzymes target low copy regions of the *P. vulgaris* genome.

SNP calling

The advances in sequencing technologies have brought advances in bioinformatics, due to the need for powerful tools for data processing. SNP calling is crucial for the correct analysis of genetic diversity and association mapping. Alignment to a reference genome works with two kinds of algorithms, the suffix tries, and the hash tables. Suffix aligners are faster and efficient with increased memory availability. The SNP calling works with two kinds of processes as well, the probabilistic and the heuristic approach. Other crucial steps in the pipeline are the removing of PCR duplications and the realignment process that corrects mismatches. Many computational programs have been created for SNP calling, and the choice of which one to use depends on the resources available and the goal of the research. A recommended pipeline includes the use of BWA (Burrows-Wheeler aligner) for alignment to the reference, SAMtools (Sequence alignment/map) for removing duplicates, and GATK (Genome analysis toolkit) for realignment of sequences (Mielczarek and Szyda, 2015).

Association mapping

Linkage disequilibrium (LD) is defined as a non-random association between two markers, two genes or quantitative trait loci (QTL), or a marker with a gene (Gupta et al., 2005). This genome characteristic can have many uses, including association mapping. Association mapping discovers linkages between a marker and a phenotype (Zhu et al., 2008). Thus, a QTL can be found using a large suite of molecular markers to discover markers associated with the

trait of interest due to a strength linkage of the molecular marker with the gene region (LD). LD can be caused by factors other than true linkage, such as population structure or relatedness within individuals. For this reason, both effects have to be controlled during the association mapping analysis (Nordborg and Weigel, 2008).

EMMA (Efficient mixed model association) is a method to determine relatedness of a population (Kang et al., 2008), while a standard principal components analysis (PCA) is often used to evaluate population structure. Considering these effects, different models are typically used to discover true associations that are not the result of structure or relatedness. A naïve model does not consider the effects of structure and relatedness; a fixed effect model can control for structure; relatedness can be treated as a random effect; and a mixed fixed and random effect model can control for both structure and relatedness. The best model can be selected based on an eye-ball test of the quantile-quantile (Q-Q) plot or the mean square deviation (Mamidi et al., 2011b), which compares observed and expected P-values of the models used to discover associations. To determine significant markers associated with a trait, several methods are used. The Bonferroni test uses the number of markers and the Type I error rate to define a cut-off for significant markers. The permutation test gives a P-value for each trait, and the cut-off values are related to the total number of markers (Mamidi et al., 2014). Since SNPs are the most abundant markers, these adapt well to the association mapping studies (Blair et al., 2013; Müller et al., 2015).

MATERIALS AND METHODS

Genetic Diversity and Population Structure Analysis

Plant material

The Institute of Agricultural Science and Technology in Guatemala (ICTA) provided two Guatemalan climbing bean populations for this analysis. The first (GUA_1966-82) was collected between years 1966 and 1982 in the western highlands of the country (Orellana et al., 2006; Ponciano-Samayoa et al., 2009) and has been maintained in the experimental center at La Alameda Chimaltenango, Guatemala. It is composed of 604 accessions belonging to three species: *P. vulgaris* (n=569), *P. dumosus* (n=18), and *P. coccineus* (n=17). For this study, 369 accessions of *P. vulgaris* were evaluated. Accessions that did not belong to *P. vulgaris* were discarded to focus only in the comparison between races of common bean. Accessions with discrepancies in color and shape within the same accession were discarded to avoid high intra accession diversity in the analysis. Passport data for this collection was lost many years ago.

Seeds of a second population (GUA_2015) were also collected in the western highlands of Guatemala in 2015 from the following departments: Chimaltenango, Quetzaltenango, Huehuetenango, San Marcos and Quiché (Figure 2). These lines were collected as part of a socio-economic survey performed by NDSU in collaboration with Michigan State University (MSU) and ICTA. The survey focused on Guatemalan smallholder farmers that grow common bean. The farmers were asked for a sample of the seed they grow on their land. This collection consists of 452 accessions belonging to three species: *P. vulgaris* (n=287), *P. dumosus* (n=138) and *P. coccineus* (n=27) (Agreda et al., 2017). For this study, 260 accessions of *P. vulgaris* were evaluated. Other accessions were excluded from the study because they belonged to different

species, had different growth habit, or because of color and shape variation within the accession. Passport data for this collection is available.

For comparative purposes, 95 race Mesoamerica (M) accessions and 95 race Durango-Jalisco (DJ) accessions were included as representatives of the Middle American gene pool. These lines are part of the Middle American diversity panel (MDP; Moghaddam et al., 2016). Additionally, 138 accessions from the Andean diversity panel (ADP; Cichy et al., 2015), including mostly race Nueva Granada accessions, were part of the study. These accessions were randomly chosen from the panels to have representatives of all the market classes belonging to these panels. Finally, 12 wild Guatemalan accessions were also included. Seeds of the wild Guatemalan accessions were provided by USDA National Plant Germplasm System. The MDP and ADP accessions were previously genotyped by the Dry Bean Genomics Lab at NDSU.

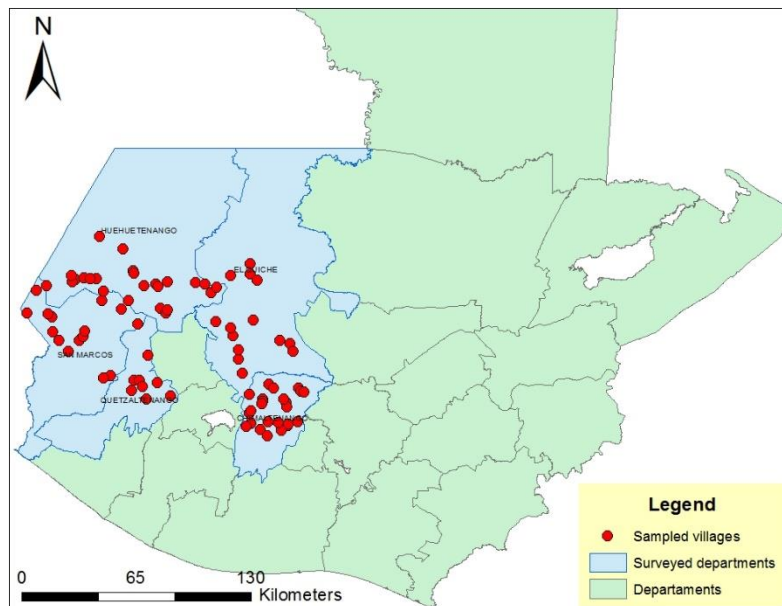


Figure 2. Map of the Guatemalan western highlands where climbing bean seed from GUA_2015 was collected. (Map source: De Young et al., 2017).

Tissue collection and DNA extraction

For the GUA_1966-82 collection and the wild Guatemalan accessions, tissue was collected from the first trifoliate leaf of two plants and placed in a 1.5 ml Eppendorf tube 15 days after planting. Tubes were immediately frozen using liquid nitrogen. For the GUA_2015 collection, leaf tissue was placed in 96 wells plates. Plates were immediately frozen using liquid nitrogen. For GUA_1966-82 (n=369 accessions), DNA isolation and extraction was done using the Genomic DNA Mini Kit (Plant) from IBI SCIENTIFIC protocol. For GUA_2015 (n=260 accessions), DNA isolation and extraction was done using the PureLink Pro 96 Genomic DNA Kit from Thermo Fisher SCIENTIFIC. DNA samples were then quantified using a NanoDrop from Thermo SCIENTIFIC and tested in a 2% agarose gel. Samples were adjusted to 20ng/ μ l.

Genotyping-by-sequencing (GBS)

DNA samples were transferred to 96 cells plates to be processed simultaneously. In this study, the optimized Genotype-by-Sequencing protocol for common bean, developed by Schröder et al. (2016) was used. 10 μ l of DNA (200 ng) and 8 units of restriction enzymes *Mse*I and *Taq*a1 were added to the plates for digesting the DNA samples in the thermocycler (37°C for 2 hours, 65°C for 2 hours, 80°C for 20 min, and hold at 8°C). Barcodes ranged from 5 to 10 base pairs (bp) and were randomly ordered in a 96 cells plate. One barcode was assigned to each digested DNA sample, and each DNA plate had the same barcode order assigned. Digested DNA, barcodes, and a ligation MasterMix were used for ligation in the thermocycler (22°C for 4 hours, 65°C for 20 min, and hold at 8°C). Small DNA fragments (<300bp) were eliminated using the AMPure Bead protocol from Agencourt, that uses magnetic beads to retain long fragments of DNA. To ensure the presence of DNA in the samples, a test PCR was performed using PCR MasterMix and primers for the barcoded adaptors (95°C for 30 sec; 36 cycles of: 95°C for 30

sec, 62°C for 20 sec, 68°C for 17 sec; 72°C for 5 min, and hold at 8°C). Results were checked using a 2% agarose gel. DNA of each plate was pooled into one library, and PCR reactions were prepared for amplification using PCR MasterMix and primers for the barcoded adaptors (95°C for 30 sec; 16 cycles of: 95°C for 30 sec, 62°C for 20 sec, 68°C for 15 sec; 72°C for 5 min, and hold at 8°C). Small DNA fragments were eliminated again with the AMPure Beads, and the library was finally validated on a 2% agarose gel to check the presence of DNA. A 100 bp ladder was used as comparison. Libraries were sent for sequencing to HudsonAlpha Institute for Biotechnology, Huntsville Alabama, USA.

Genotypes separation and barcodes trimming

Sequence data was received as compressed FASTQ files (one file per plate). A genotype file was created that included the genotype name, barcodes information for each genotype, plate and well assigned to each genotype, read group ID (RGID), read group library identifier (RGLB), read group platform (RGPL), and read group platform unit (RGPU). The last four were provided by HudsonAlpha. Using the barcode information, sequences for each genotype were separated and placed in separate folders using the FASTX toolkit barcode splitter (http://hannonlab.cshl.edu/fastx_toolkit/). Once the files were separated based on the barcodes and given unique names, the barcodes were trimmed from the sequences using the FASTA/Q trimmer from FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/).

Quality control

To check the quality of the sequences, the software SICKLE from UC Davis (Joshi and Fass, 2011) was used. The software discarded all the reads that were less than 80bp length. The default quality threshold was 20; if quality was too low, the 3' end of the reads were trimmed

and if quality was high enough the 5' end of the reads were trimmed, considering the assumption that sequencing can produce reads with low quality towards the 3' and 5' ends of the reads.

Mapping and read group information

The Burrows-Wheeler Alignment (BWA) Tool (Li, 2013) was used to map the reads to the reference genome. Version 2 assembly of Race Peru landrace G19833 was used as the reference (Schmutz et al., 2014; https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Pvulgaris). The main genome (11 chromosomes) was indexed using BWA and SAMtools faidx (Li et al., 2009) to generate the BAM file for each genotype. At this step, the MDP and ADP genotypes were included in the analysis. Since read group information is necessary for SNP calling, the information (RGID, RGLB, RGPL and RGPU) was added to the BAM files using Picard tools (<http://broadinstitute.github.io/picard>).

SNP calling

Variant calling was performed using the Unified Genotyper from the Genome Analysis Toolkit (GATK) (DePristo et al., 2011). The standard minimum confidence threshold for calling was 30. Eleven VCF files were created (one for each chromosome). VCF files were then filtered for read depth with GATK, and any call with less than three reads was discarded. The filtered VCF files were filtered again for the next criteria: at least five genotypes should have the reference allele, at least five genotypes should have the alternative allele, and there has to be less than 50% missing data. All the chromosomes were merged to a single file using the CAT function in Linux, and the header was added manually.

Data imputation

Missing data was imputed using the software fastPHASE in Linux (Scheet and Stephens, 2006). The VCF file was first converted to a HAPMAP file using Tassel 5 (Bradbury et al.,

2007). Markers with more than two alleles and genotypes with more than 90% missing data were discarded using R (R Core Team, 2015; www.R-project.org). Missing data were converted to question marks and base pairs were converted to “AA” for homozygous reference alleles, “BB” for homozygous alternative alleles, and “AB” for heterozygous alleles using a Perl function (<http://strawberryperl.com>), and the file was then transposed. The file was converted to the fastPHASE format using R. In the fastPHASE command line, the lower limit for number of clusters (KL) was set at 2, the upper limit for number of clusters (KU) was set at 40, and the interval between values for number of clusters (Ki) was set at 2. The imputed file was converted again to a HAPMAP file with its original alleles using a Perl function and R. Using TASSEL 5, the file was filtered for minor allele frequency (MAF) of 5% and for no more than 50% of heterozygous markers. The pipeline used for SNP calling is described by Mielczarek and Szyda (2015).

Population structure analysis

STRUCTURE

Alleles in the HAPMAP file were converted to numbers (G=1, C=2, A=3, T=4). STRUCTURE, version 2.3 (Pritchard et al., 2000) was used to estimate the number of subpopulations (K) represented by the population. An “Admixture Model with Allele Frequencies Correlated among Populations” was used with 10 runs and with K size ranging from 1 to 10. Burnin period and number of iterations were set at 20000/10000, based on McClean et al. (2012) in their comparison of two burnin/iteration combinations. Evano’s ΔK statistic (Evano et al., 2005) was used to determine the most likely number of subpopulations. The software DISTRUCT (Rosemberg, 2004) was used to display the results in a graph.

Principal components analysis (PCA)

PCA was performed using R (R Core Team, 2015; www.R-project.org) with the `prcomp` function. Alleles in the HAPMAP file were converted to numeric data, where “0” was homozygous reference, “2” was homozygous alternative, and “1” was heterozygous. The principal components matrix, variance explained by each principal component, and the cumulative variance were calculated. PCA plots for PC1 vs PC2, PC2 vs PC3 and PC1 vs PC3 were created using the `ggbiplot` function (<http://github.com/vqv/ggbiplot>) in R.

Maximum likelihood (ML) tree

The ML tree was developed using the software SNPhylo (Lee et al., 2014), which accepts the HAPMAP file. Linkage disequilibrium (LD) was set at 0.2 and the unrooted tree was bootstrapped 1000 times. The software creates a FASTA file with a subset of markers in LD that are used to create the tree file and can also be used with other software like MEGA (Kumar et al., 2016). The resulting tree file was opened with Figtree (<http://tree.bio.ed.ac.uk/software/figtree/>) for visualization. These results were compared with the PCA and the STRUCTURE analyzes.

Genetic diversity analysis

To calculate genetic diversity statistics, the original allele data of the HAPMAP were used. Genetic statistics such as observed heterozygosity (H_o), Polymorphic Information Content (PIC), and Nei’s expected heterozygosity (H_e) (Nei and Li, 1979), were estimated using the software PowerMarker (Liu and Muse, 2005). A pairwise population differentiation (F_{ST}) matrix and an analysis of molecular variance (AMOVA) for gene pools, races of the Middle American gene pool and Guatemalan collections were generated using the software GenAlex (Peakall and Smouse, 2006, 2012). ΔH (Vigouroux et al., 2002) was used to determine the loss of genetic diversity of the climbing bean collections when compared to the wild accessions. $\Delta H = 1 -$

(H_c/H_w), where H_c and H_w represent the expected heterozygosity of the climbing bean accessions and the Guatemalan wild accessions, respectively.

Geographical distribution in the GUA_2015 collection

The association between genetic structure and geography of the second population of climbing beans was analyzed using PCA. Genotypes were classified first based on its range of elevation (meters above the sea level) and then by location (department where seed was collected), and a different color was assigned to each group. PCA was performed using R (R Core Team, 2015; www.R-project.org) with the `prcomp` function. Alleles in the HAPMAP file were converted to numeric data, where “0” was homozygous reference, “2” was homozygous alternative, and “1” was heterozygous. The principal components matrix, variance explained by each principal component, and the cumulative variance were calculated. PCA plots for PC1 vs PC2, PC2 vs PC3 and PC1 vs PC3 were created using the `ggbiplot` function (<http://github.com/vqv/ggbiplot>) in R.

Intra-accession diversity analysis

Population

To evaluate the genetic diversity within accessions in the GUA_1966-82 collection, a sample of 48 accessions was used, including accessions of the three species: *P. vulgaris* (n=40), *P. dumosus* (n=4), and *P. coccineous* (n=4). Accessions for this analysis were randomly chosen from the original GUA_1966-82 collection (604 accessions). The second collection was not included in this analysis due to a limited seed supply.

Tissue collection and DNA extraction

12 seeds of each accession were planted in trays of eighteen cells at the NDSU greenhouse. Tissue was collected 15 days after planting from the first trifoliate leaf of each plant

and placed in 96 wells plates (12 columns x 8 rows), where each row in the plate represented one genotype. Plates were immediately frozen using liquid nitrogen. DNA isolation and extraction was done using the PureLink Pro 96 Genomic DNA Kit from Thermo Fisher SCIENTIFIC. DNA samples were then quantified using NanoDrop from Thermo SCIENTIFIC.

Polymerase chain reaction (PCR) and electrophoresis

A set of 33 insertion-deletion (InDel) markers (three for each chromosome) for black beans were chosen from the set of market class specific InDel markers developed by Moghaddam et al. (2014). The sequences of the forward and reverse primers for the InDel markers, as well as their physical positions are listed in Table A1. DNA was amplified in the thermocycler running at 94°C during 3 min; 45 cycles of 94°C during 20 sec, 55°C during 30 sec, and 72°C during 1 min; a run at 72°C during 10 min and held at 12°C. Amplified DNA was run in a 3% agarose gel with 7 µl of ethidium bromide. Each gel contained 288 samples. Gels were then photographed in an AlphaImager HP system from ProteinSimple. Bands for each marker were then quantified and converted to numeric data, where each allele represented a number. Numeric file was used for diversity analysis.

Genetic diversity analysis

Heterozygosity (H_o), polymorphic information content (PIC), and Nei's expected heterozygosity (H_e) were estimated for each genotype using the software PowerMarker (Liu and Muse, 2005). General H_e was also calculated. The proportion of the genetic variation found among accessions and within accessions was also calculated using AMOVA implemented in GenAlex (Peakall and Smouse, 2006, 2012).

Genome-Wide Association Study (GWAS)

SNP markers were compared with phenotypic data to discover important genetic associations. Phenotypic data for the analysis was described by Orellana et al. (2006) for the GUA_1966-82 collection, and the traits analyzed are listed on Table 1. Elevation data was used as the dependent variable for the GUA_2015 collection. This data was provided by the Agriculture, Food and Resources Department at Michigan State University. Other agromorphological traits are not available for GUA_2015.

The only trait measured at NDSU Dry Bean Genomics lab was seed shape. Seed shape was measured in the genotypes of the GUA_1966-82 population. The length (mm), and width (mm) of five seeds for each genotype were measured with a digital caliper Performance Tool. Least Square Means (LSMeans) of length, width, and the ratio of width/length was calculated in SAS version 9.3 (SAS Institute Inc. 2011). In the case of disease resistance, a scale from 1 (resistant) to 9 (susceptible) was used. However, the measure was based on natural pressure of the diseases in the place where the plants were grown, and no specific races of pathogens were evaluated.

GAPIT package (Zhang et al., 2010) in R (R Core Team, 2015; www.R-project.org) was used to perform the GWAS for most of the traits. For traits scored as binary data, a logistic regression analysis was performed using the R package GENABEL (Aulchenko et al., 2007). Several models were tested in the study: 1) naïve linear model; 2) structure fixed effect control model; 3) relatedness random effect control model; and 4) structure and relatedness mixed effect control model. The best model was selected for each trait based on the calculation of the mean square deviation (MSD) (Mamidi et al., 2011b) and the quantile-quantile plots. Significant SNP markers were defined using a permutation test with 10,000 permutations and a cut-off p-value

was set as lowest at 0.1 percentile (Mamidi et al., 2014). Population structure was controlled using a PCA matrix, and relatedness was controlled using a Kinship matrix that was calculated with the EMMA algorithm in GAPIT. R-square values of the peak SNP for each trait, as well as R-square values for the most important SNPs associated with each trait were calculated in R using GENABEL, which accounts for structure and relatedness during calculation. Candidate genes were searched based on 100 kb upstream and downstream window relative to the location of the peak SNPs (Moghaddam et al., 2016) in Version 2 of the genome annotation of *P. vulgaris* (Schmutz et al., 2014; https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Pvulgaris).

Table 1. Phenotypic data for the GUA_1966-82 population.

| Feature | Scale |
|-------------------------------------|----------------------------------|
| Altitudinal adaptation ¹ | masl ² |
| Cotyledon color | 1 to 8 |
| Flower color | 1 to 7 |
| Growth habit | Type I, II, III or IV. |
| Stem color | 1 to 5 |
| Hypocotyl color | 1 to 6 |
| Pod color | 1 to 9 |
| Pod shape | 1 to 4 |
| Pod distribution in the plant | 1 to 4 |
| Pod profile | 1 to 4 |
| Seed color | 1 to 18 |
| Seed coat luster | Dull or shiny |
| Seed length | mm |
| Seed width | mm |
| Seed length/width ratio | - |
| Rust ³ | 1 (resistant) to 9 (susceptible) |
| Anthraco nose | 1 (resistant) to 9 (susceptible) |
| Aschochyta | 1 (resistant) to 9 (susceptible) |
| Phytophthora | 1 (resistant) to 9 (susceptible) |
| Angular leaf spot | 1 (resistant) to 9 (susceptible) |

The scales follow the methodology of the varietal descriptors developed by the International Center for Tropical Agriculture (CIAT) for common bean (Muñoz et al., 1993). ¹Altitudinal adaptation was the only trait that was evaluated in the GUA_2015 collection. ²masl=meters above sea level. ³Disease resistance scores are based on natural pressure of the pathogen because specific races were not available at the time of evaluation.

RESULTS

Genetic Diversity and Population Differentiation

SNP calling

A total of 971 *P. vulgaris* genotypes were analyzed including the two Guatemalan collections, the Guatemalan wild accessions, members of races Mesoamerica and Durango-Jalisco, and Andean gene pool accessions. A total of 134,152 SNPs were discovered for all the Middle American races including the Guatemalan groups. The ADP collection consisted of 265,765 SNPs. A total of 45,128 shared ADP and Middle American SNPs, distributed across all 11 linkage groups of common bean, were used for the analysis.

Genetic diversity analysis

For the genetic diversity analysis of the 971 genotypes, the 2,453 SNPs with local LD values less than 0.2 were evaluated. 100% of the markers were polymorphic over the entire population. However, when they were separated by subpopulations, the percentage of polymorphic markers ranged from 94.8% (wild accessions subpopulation) to 100% (Middle American gene pool). All markers had two alleles per locus. Mean diversity statistics for each subpopulation are presented in Table 2.

When the overall population was evaluated, the mean H_e was 0.41 with a range of 0.18 to 0.50. The lowest H_o was 0.003, and the highest was 0.96, with a mean of 0.32. The lowest PIC value was 0.16 and the highest was 0.38, with a mean of 0.32. When the two gene pools were compared H_e , H_o and PIC values were higher in the Middle American gene pool ($H_e=0.38$; $H_o=0.34$; PIC=0.30) than in the Andean gene pool ($H_e=0.24$; $H_o=0.17$; PIC=0.20). The H_e for the Guatemalan wild subpopulation (0.35) was higher than H_e for the GUA_1966-82 (0.31). However, it was lower when compared to H_e in GUA_2015 (0.36). Consequently, diversity loss

(ΔH) was 11% in GUA_1966-82, while there was almost 3% of diversity gain in GUA_2015 relative to the wild population.

In general, the greatest diversity difference was between the two gene pools, while the magnitude of diversity was similar among the different Middle American subpopulations. The Andean gene pool had the lowest H_e among all subpopulations (0.24), while GUA_2015 had the highest H_e (0.36). Among the Middle American groups, GUA_1966-82 had the lowest H_e (0.31), followed by M (0.34), DJ (0.35) and GUA_2015 (0.36). The group DJ had the highest H_o (0.36) and GUA_1966-82 had the lowest (0.32). As expected, GUA_2015 had the highest PIC value (0.28), followed by DJ (0.27), M (0.26) and GUA_1966-82 (0.25).

Table 2. Summary of diversity statistics of the different populations.

| Subpopulation | Genotypes (n) | % Polymorphic SNPs | PIC | | | |
|-------------------|------------------|-----------------------|-------|-------|--------|------------|
| | | | H_e | H_o | values | ΔH |
| MA gene pool | 819 | 100.00 | 0.38 | 0.34 | 0.30 | - |
| GUA-1966-72 | 369 | 99.28 | 0.31 | 0.32 | 0.25 | 0.11 |
| GUA-2015 | 260 | 99.83 | 0.36 | 0.35 | 0.28 | -0.03 |
| Mesoamerica | 95 | 97.27 | 0.34 | 0.35 | 0.26 | - |
| Durango-Jalisco | 95 | 96.87 | 0.35 | 0.36 | 0.27 | - |
| Guatemalan wild | 12 | 94.83 | 0.35 | 0.31 | 0.28 | - |
| Andean gene pool | 138 | 99.93 | 0.24 | 0.17 | 0.20 | - |
| Entire population | 971 | 100.00 | 0.41 | 0.31 | 0.32 | - |

MA=Middle American; H_e =Expected heterozygosity; H_o =Observed heterozygosity; PIC=polyomorphic information content; ΔH =Diversity loss, when ΔH is negative, the diversity of the domesticated population is higher than the diversity of the wild population.

Intra-accession diversity analysis

A total of 33 InDel markers and 48 accessions from GUA_1966-82 were used to evaluate diversity among individuals within an accession. Among all the markers, four were monomorphic across all accessions. The remaining 29 markers had 2 alleles. The percentage of monomorphic markers for the accessions ranged from 12.12 (Guate248) to 96.97 (Guate200). H_e ranged from 0.013 (Guate1003) to 0.26 (Guate248) with a mean of 0.09 (Table 6). H_o ranged

from 0.0 (Guate200) to 0.12 (Guate1112) with a mean of 0.04. PIC values ranged from 0.01 (Guate1003) to 0.21 (Guate248) with a mean of 0.07. In general, just two accessions had H_o higher than 0.1, which means that the heterozygosity within all of the accessions is low. In contrast, almost half of the accessions evaluated had H_e higher than 0.1, which means that diversity within accessions is still high. Results of the AMOVA (Table 3) showed that variation among accessions (59%) is greater than variation within accessions (41%). However, within accession diversity is still considered high.

Table 3. Analysis of molecular variance (AMOVA) among and within accessions.

| Source | df | MS | St. Var. | Variation explained | P-value |
|-------------------|------|-------|----------|---------------------|---------|
| Among accessions | 47 | 82.17 | 3.33 | 59% | <0.001 |
| Within accessions | 1104 | 2.28 | 2.28 | 41% | <0.001 |
| Total | 1151 | | 5.61 | 100% | |

df = degrees of freedom.. MS = mean squares. St. Var. = Standard variation.

Population differentiation

All pairwise F_{ST} values among the six subpopulations evaluated were significant ($\alpha=0.05$). The degree of differentiation between any two subpopulations ranged from 0.026 to 0.264 (Table 4). All the Middle American races including race Guatemala had a high genetic differentiation when compared with the Andean gene pool. The genetic differentiation within the Middle American subpopulations was moderate. It was interesting to note that differentiation between DJ and GUA_2015 was lower than between M and GUA_2015. The highest differentiated subpopulations were race M and the Andean group (0.264). The two subpopulations that had lowest genetic differentiation (0.026) were GUA_1966-82 and GUA_2015. Interestingly, the Guatemalan wild subpopulation was highly differentiated from the Andean, DJ and M groups, while it was just moderately differentiated with the two Guatemalan subpopulations.

Table 4. Pairwise F_{ST} values of all the populations evaluated.

| | Andean gene pool | Durango/ Jalisco | Mesoamerica | GUA_2015 | GUA_1966-82 | Guatemalan wild |
|-------------------------|---------------------------------|-----------------------------|--------------------|-----------------|--------------------|----------------------------|
| Andean gene pool | 0.000 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| Durango/Jalisco | 0.240 | 0.000 | 0.001 | 0.001 | 0.001 | 0.001 |
| Mesoamerica | 0.264 | 0.073 | 0.000 | 0.001 | 0.001 | 0.001 |
| GUA_2015 | 0.195 | 0.099 | 0.134 | 0.000 | 0.001 | 0.001 |
| GUA_1966-82 | 0.217 | 0.125 | 0.161 | 0.026 | 0.000 | 0.001 |
| Guatemalan wild | 0.155 | 0.147 | 0.157 | 0.084 | 0.109 | 0.000 |

F_{ST} = Fixation index. F_{ST} values are below the diagonal. Probability ($\alpha=0.05$) based on 1000 permutations is shown above the diagonal.

AMOVAs were performed to evaluate the variation explained by gene pools, races of the Middle American gene pool and, the Guatemalan collections. Table 5 showed that variation among gene pools (29%) is lower than variation explained within gene pools (71%). Similarly, the variation among races (22%) is lower than variation within races (78%). In general, the variation within each gene pool is higher than the variation between gene pools. On the other hand, variation among the Guatemalan groups is very low (5%) and most of the variation is explained within the groups. These results are in accordance with the results of the pairwise F_{ST} that showed that the less differentiated subpopulations are the Guatemalan groups, and the most differentiated subpopulations are the Andean and Middle American gene pools.

Table 5. AMOVA for separation of gene pools, Middle American races and Guatemalan groups.

| Source of variation | df | MS | St. Var. | Variation explained | P values |
|----------------------------|-----------|-----------|-----------------|----------------------------|-----------------|
| Gene pools | | | | | |
| Among gene pools | 1 | 86241.16 | 181.61 | 29% | <0.001 |
| Within gene pools | 1912 | 448.19 | 448.19 | 71% | <0.001 |
| Total for gene pools | 1913 | | 629.80 | 100% | |
| Races | | | | | |
| Among races | 2 | 37024.08 | 116.61 | 22% | <0.001 |
| Within races | 1635 | 421.47 | 421.47 | 78% | <0.001 |
| Total for races | 1637 | | 538.08 | 100% | |
| Guatemalan groups | | | | | |
| Among groups | 1 | 13657.74 | 21.72 | 5% | <0.001 |
| Within groups | 1256 | 408.58 | 408.58 | 95% | <0.001 |
| Total for groups | 1257 | | 430.30 | 100% | |

df = degrees of freedom. MS = mean squares. St. Var. = Standard variation.

Table 6. Summary of intra-accession diversity statistics for GUA_1966-82 genotypes.

| Genotype | H_e | H_o | PIC | % monomorphic markers |
|-----------------|-------------------------|-------------------------|------------|------------------------------|
| Guate248 | 0.2641 | 0.0397 | 0.2118 | 12.12 |
| Guate192 | 0.0481 | 0.0341 | 0.0442 | 90.90 |
| Guate200 | 0.0142 | 0.0000 | 0.0109 | 96.97 |
| Guate233 | 0.0373 | 0.0278 | 0.0319 | 81.82 |
| Guate891 | 0.0393 | 0.0303 | 0.0313 | 87.88 |
| Guate949 | 0.0328 | 0.0303 | 0.0272 | 87.88 |
| Guate977 | 0.1197 | 0.0588 | 0.0946 | 69.70 |
| Guate1246 | 0.1650 | 0.0732 | 0.1360 | 45.45 |
| Guate135 | 0.0964 | 0.0135 | 0.0815 | 63.64 |
| Guate385 | 0.0686 | 0.0000 | 0.0589 | 69.70 |
| Guate578 | 0.1661 | 0.0599 | 0.1327 | 51.52 |
| Guate992 | 0.0471 | 0.0135 | 0.0381 | 81.82 |
| Guate1000 | 0.0909 | 0.0323 | 0.0833 | 87.88 |
| Guate1242 | 0.0569 | 0.0462 | 0.0479 | 69.70 |
| Guate297 | 0.0764 | 0.0283 | 0.0625 | 78.79 |
| Guate639 | 0.0948 | 0.0340 | 0.0749 | 72.73 |
| Guate888 | 0.1388 | 0.0320 | 0.1110 | 60.61 |
| c815cm6 | 0.0785 | 0.0193 | 0.0642 | 75.76 |
| Guate10071 | 0.1578 | 0.1047 | 0.1291 | 48.48 |
| Guate1069 | 0.0182 | 0.0061 | 0.0150 | 93.94 |
| Guate1222 | 0.1054 | 0.0495 | 0.0877 | 63.64 |
| Guate1234 | 0.0767 | 0.0177 | 0.0634 | 72.73 |
| Guate945 | 0.0847 | 0.0269 | 0.0689 | 72.73 |
| Guate952 | 0.0436 | 0.0344 | 0.0376 | 78.79 |
| Guate1005 | 0.1482 | 0.0636 | 0.1188 | 57.58 |
| Guate1100 | 0.0653 | 0.0076 | 0.0561 | 75.76 |
| Guate1244 | 0.0606 | 0.0313 | 0.0530 | 90.91 |
| Guate12803PM | 0.0814 | 0.0278 | 0.0696 | 66.67 |
| LCH86V87 | 0.1012 | 0.0407 | 0.0850 | 60.61 |
| Guate1053 | 0.0752 | 0.0379 | 0.0649 | 63.64 |
| Guate1073 | 0.1017 | 0.0396 | 0.0820 | 72.73 |
| Guate1511 | 0.0405 | 0.0076 | 0.0342 | 84.85 |
| LaborOvalle | 0.0764 | 0.0404 | 0.0645 | 72.73 |
| LCH86V23 | 0.1817 | 0.0404 | 0.1428 | 54.55 |
| c495cm6cm12 | 0.0992 | 0.0583 | 0.0795 | 69.70 |
| c495cm6cmb | 0.0806 | 0.0185 | 0.0662 | 75.76 |
| Guate1112 | 0.1856 | 0.1200 | 0.1550 | 24.24 |
| Guate1418 | 0.1290 | 0.0383 | 0.1038 | 63.64 |
| Guate1429 | 0.1101 | 0.0482 | 0.0927 | 60.61 |

H_e =Expected heterozygosity; H_o =Observed heterozygosity; PIC=polymorphic information content.

Table 6. Summary of intra-accession diversity statistics for GUA_1966-82 genotypes (continued).

| Genotype | H_e | H_o | PIC | % monomorphic markers |
|-----------------|-------------------------|-------------------------|------------|------------------------------|
| Guate1003 | 0.0125 | 0.0177 | 0.0099 | 96.97 |
| Guate1192 | 0.1548 | 0.0863 | 0.1249 | 48.48 |
| Guate1275 | 0.0369 | 0.0078 | 0.0362 | 93.94 |
| Guate1451 | 0.0814 | 0.0573 | 0.0710 | 81.82 |
| Guate1470 | 0.0843 | 0.0708 | 0.0741 | 78.79 |
| Guate1101 | 0.0417 | 0.0104 | 0.0395 | 93.94 |
| Guate1210 | 0.0797 | 0.0717 | 0.0701 | 81.82 |
| Guate1286 | 0.0349 | 0.0052 | 0.0346 | 93.94 |
| Guate1313 | 0.0387 | 0.0052 | 0.0376 | 93.94 |
| MEAN | 0.0880 | 0.0368 | 0.0731 | 72.35 |

H_e =Expected heterozygosity; H_o =Observed heterozygosity; PIC=polymorphic information content.

Population structure analysis

The population structure analysis was performed using 971 genotypes and 2,453 SNPs (local LD <0.2). Three subpopulations (K=3) were identified as the most important hierarchical level of structure using ΔK (Figure 3) based on the Evanno et al. (2005) calculation. The first partition (K=2) split the population into Middle American and Andean gene pools. When K=3, the Guatemalan collections (GUA) separated from races M and DJ (MA group). K=4 split race M from race DJ. At K=5, GUA_1966-82 separated from GUA_2015. Wild accessions never formed a separate group. To assign genotypes to each subpopulation (using K=3), genotypes with $q_i < 0.70$ were classified as admixed. A single accession from GUA_2015 was classified as Andean. Ten accessions classified originally as Andean were assigned to the MA group. Also four accessions classified originally as GUA were assigned to the MA group. Four accessions of the Andean group were admixed with the MA group. 16 accessions of the MA group were admixed with the GUA group. Additionally, 46 accessions from the GUA group were admixed with the MA group. Within the wild accessions, one was assigned to the MA group, two were

assigned to the GUA group and nine were admixed between the MA group, the GUA group and the Andean gene pool.

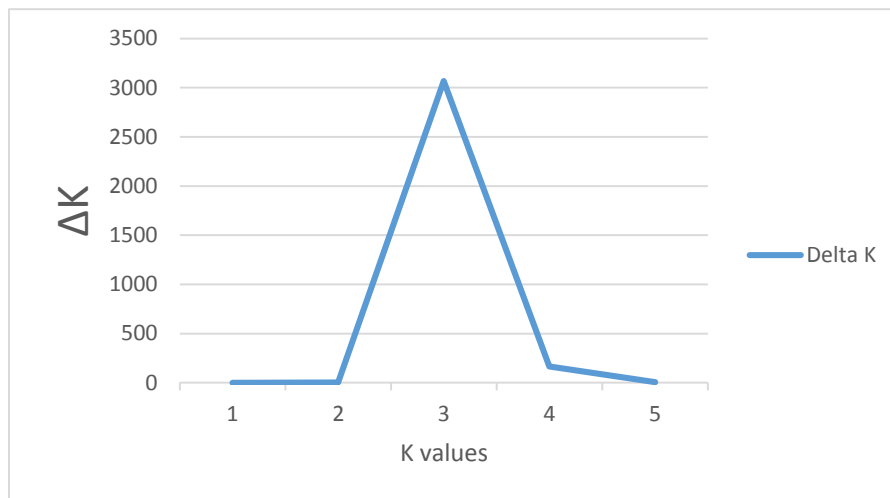


Figure 3. ΔK results showing the best $K=3$.

$K=4$ was also analyzed to evaluate races M and DJ separately. Two accessions classified originally as M were assigned to the DJ group, and six accessions classified originally as DJ were assigned to the M group. Five genotypes of the M group were admixed with the DJ group and 13 genotypes of the DJ group were admixed with the M group. In general, the DJ group had more admixture (37%) compared to M group (14%), and most of this admixture included the GUA group. Interestingly, with $K=4$ it was revealed that wild accessions had admixture of all the groups except DJ.

With $K=5$, GUA_1966-82 and GUA_2015 split. However, GUA_2015 had a high percentage of admixture (84%), with most of it explained by admixture of the two Guatemalan groups. Interestingly, 49 genotypes from GUA_2015 were admixed with the DJ group. In contrast, just six genotypes were admixed with the M group. Just 17 genotypes (5%) of GUA_1966-82 were admixed, most of them with the M group. Figure 4 shows the results for population structure analysis using STRUCTURE.

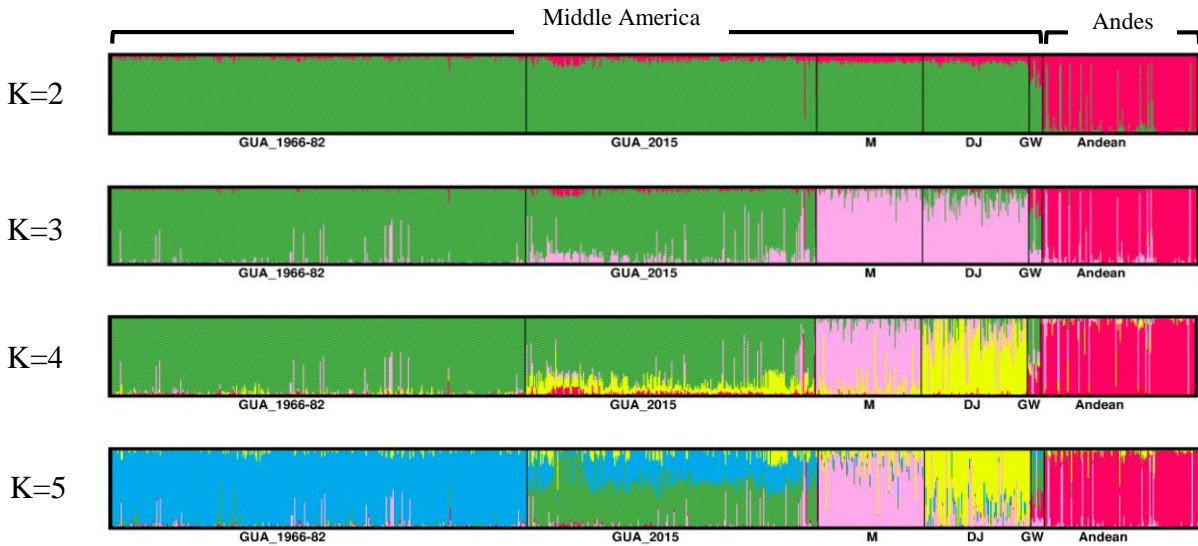


Figure 4. Results of population structure analysis with K values ranging from 2 to 5. M= race Mesoamerica, DJ= race Durango_Jalisco, GW= Guatemalan wilds.

Principal components analysis

PCA analysis was performed using 971 genotypes and all 45,128 SNPs markers. This analysis was consistent with the STRUCTURE results. The first component explained 22.1% of the variation and separated the Andean gene pool from the Middle American gene pool. The second component explained 16.9% of the variation and separated the Guatemalan collections from the Middle American races. The third component explained 4.9% of variation and separated race M from race DJ (Figure 5). In the plots, it was also observed that one accession from GUA_2015 clustered with the Andean gene pool. Nine accessions classified as Andean clustered with the M group, and one clustered with the DJ group. One wild accession was clustered in the M group. Admixture between the two Guatemalan collections and between races M and DJ was also observed in the PC plots. These misclassifications and admixtures were also identified by STRUCTURE. Interestingly, admixture between some GUA_2015 accessions with the DJ group is observed in the plots. The STRUCTURE results showed that a higher percentage of GUA_2015 were admixed with the DJ group. In general, high admixture between Middle

American gene pool races is observed. In contrast, low admixture between gene pools is detected.

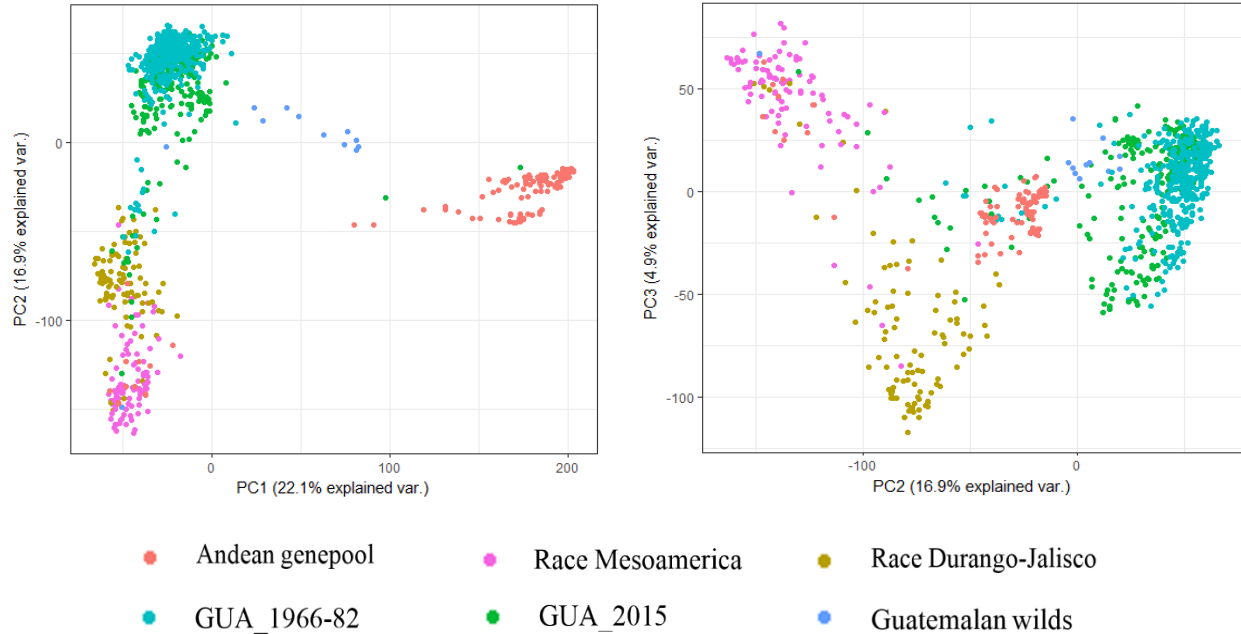


Figure 5. Results of PCA for the first three principal components based on the SNPs information.

Phylogenetic analysis

A maximum likelihood tree was generated using 971 genotypes and 2,453 SNPs (local LD<0.2). The results matched with the analyses using STRUCTURE and principal components. A cluster formed by the lower branches of the tree was represented mostly by accessions from the Andean gene pool, while a cluster formed by the upper branches of the tree was represented by accessions of races M and DJ. The Guatemalan collections clustered mostly together but it was observed that they formed subgroups. It was also observed that GUA_2015 may be a subpopulation of GUA_1966-82 since many accessions of this subpopulation are interleaved within the GUA_1966-82 subpopulation. The tree also detected this relationship. One wild accession clustered with the M group, 12 accessions classified as Andeans clustered with the M group and one with the DJ group. Five accessions classified as M clustered with the group DJ,

and 10 classified as DJ clustered with the M group. Two accessions from GUA_2015 were clustered with the Andean group and one with the MA group. Guatemalan accessions that were admixed with the M group in STRUCTURE clustered together closer to the M group in the tree. Guatemalan accessions that had admixture with the DJ group clustered together in one of the Guatemalan sub clusters. Figure 6 shows results of the ML tree with the branches collapsed. The original tree is showed in Figure A1.

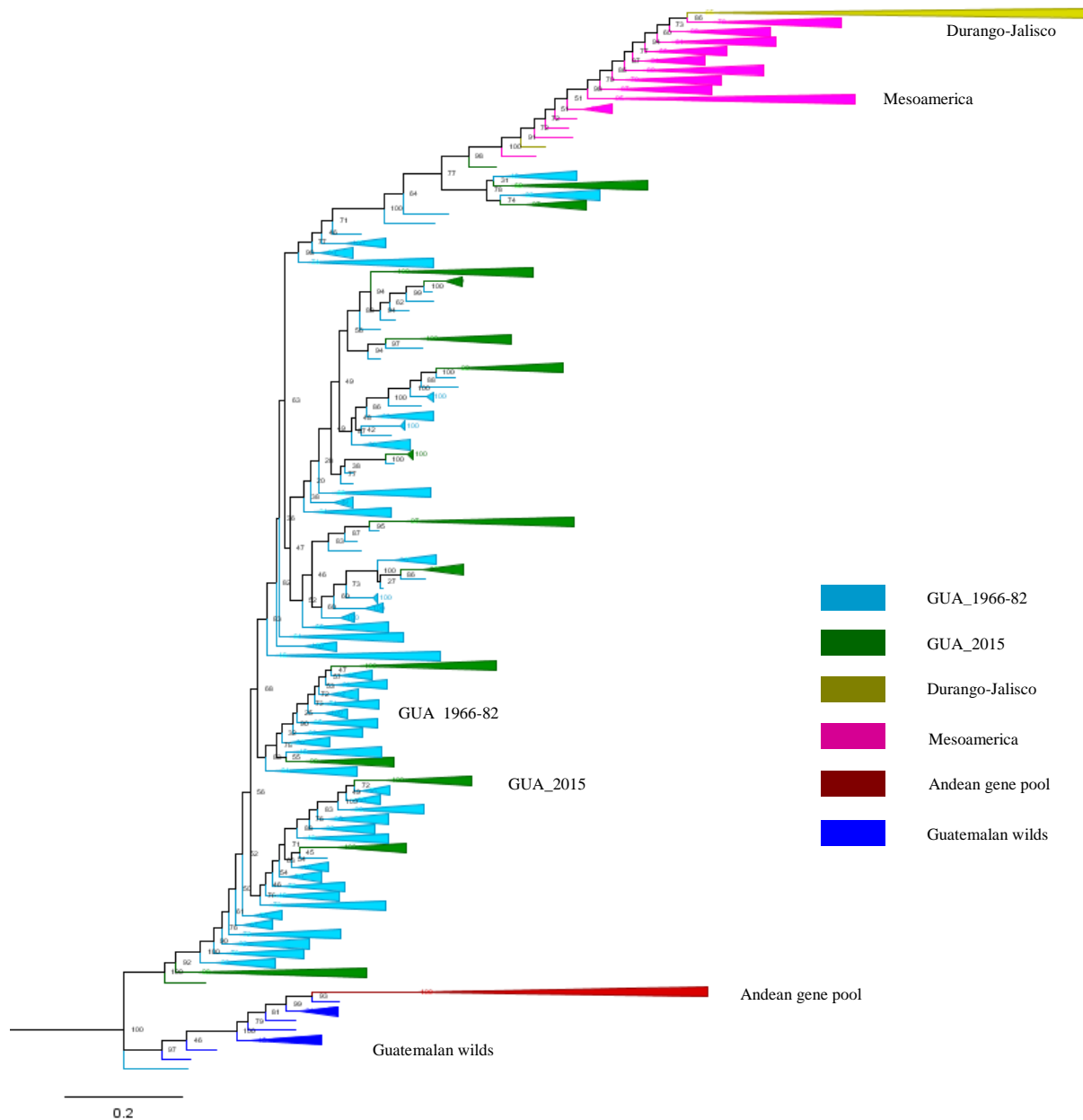


Figure 6. Results of the maximum likelihood tree. Clustered branches are collapsed.

Geographical structure of the genetic variation in population GUA_2015

This analysis was performed only with population GUA_2015 because no passport data was available for population GUA_1966-82. SNP calling for the GUA_2015 collection was performed separately. A total of 112,592 SNPs were discovered originally and the total of SNPs after filtering for a 5% MAF was 104,110. SNP loss after filtering was 7.5%. PCA was performed using all the SNPs and 253 genotypes. Accessions that did not belong strictly to the type IV climbing growth habit were discarded. Genotypes were identified based on the collection location and the elevation of those regions, and PCA analysis was performed individually for each, location and elevation data. PCA plots showed that PC1 explained 11.8% and PC2 explained 9.5% of the total genetic variation. Figure 6a showed that PC1 separated accessions collected in Huehuetenango from the rest of locations, and PC2 separated most of the accessions collected in Chimaltenango from accessions collected in Quetzaltenango. Accessions from Quiche and San Marcos do not form a specific cluster. Interestingly, when genotypes were identified with elevation, PC1 separated those genotypes grown at less than 1500 masl from genotypes grown at more than 2500 masl. Genotypes grown in intermediate elevation were mostly in the middle of the plot. In general, the elevation of the locations where the accessions were grown is an important factor for population structure of GUA_2015.

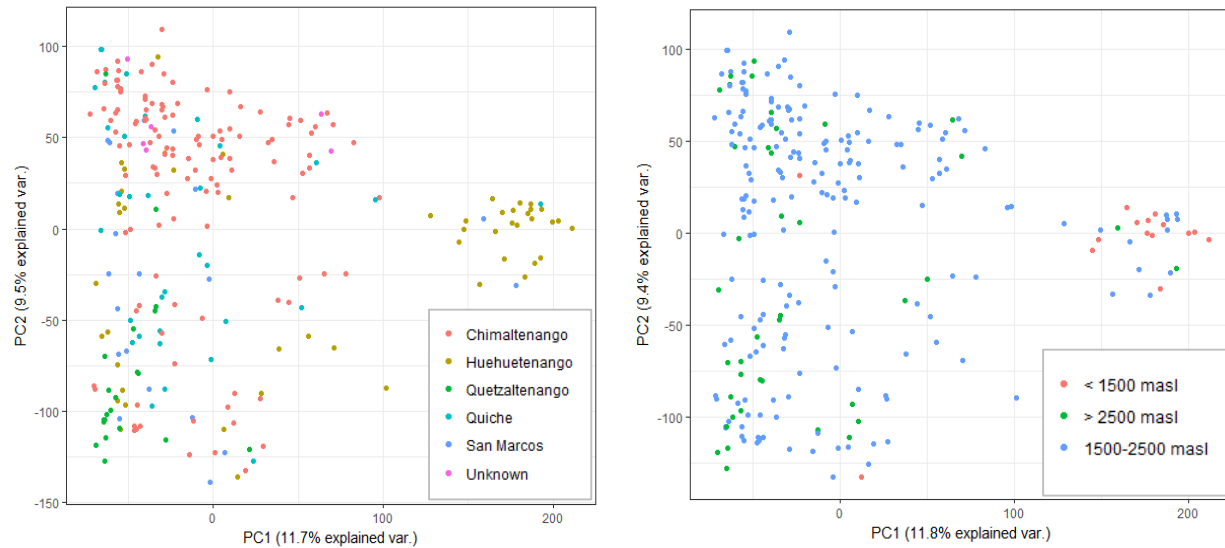


Figure 7. Results of PCA for the first two principal components.

Genome Wide Association Study

369 accessions from GUA_1966-82 and 78,754 SNPs were used for the GWAS analysis. SNP calling of only GUA_1966-82 identified a total of 102,822 SNPs. SNP loss after filtering for MAF and heterozygosity (0.05 and 0.5, respectively) was 23%. Phenotypic data were provided by ICTA, and a description of the scales used to score the traits are provided in Table 2. Since nine traits did not present strong associations with genomic regions, only 12 of the 20 traits evaluated are analyzed in this section. GWAS results for the other nine traits are presented in the Figure A2.

Altitudinal adaptation was the only trait that was evaluated exclusively for GUA_2015. GWAS analysis for this population was performed using its own set of markers (104,110 SNPs) and 256 genotypes (some of them lack data for elevation), and it used 3PCs (25% of variation) to control for population structure instead of 6PCs that was used for the other traits.

The best GWAS models were chosen based on the lowest MSD values (Mamidi et al., 2014) and the best quantile-quantile (Q-Q) plots. Table 7 shows the MSD values for each model

and each trait evaluated. Q-Q plots show how the data is adjusted relative to the chosen models (Figures 8b to 15b). For most of the traits, Q-Q plots and MSD values agreed. Also, mixed models usually gave the lowest MSD values, while the naïve models was not the best fitting model for any trait. This demonstrates the strong influence of relatedness and population structure for the expression of the traits. In the case of seed coat luster and rust resistance, the best model was chosen based only on the Q-Q plots. Many genomic regions were associated with the traits evaluated. Table A2. lists the significant markers for the 0.01 percentile and the candidate genes based on an evaluation of 100 kb downstream and upstream window relative to the marker position for each trait. Manhattan plots (Figures 8a to 15a) show the significant markers, with chromosomal position on the x-axis and the $-\log_{10}(P)$ values on the y-axis.

Table 7. Mean Square Deviation (MSD) for each GWAS model and each trait evaluated.

| Trait | Mixed Model ² | EMMA ³ | 6PC ⁴ | Naïve |
|-------------------------------------|--------------------------|-------------------|------------------|----------|
| Altitudinal adaptation ¹ | 3.85E-05 | 1.21E-55 | 9.52E-03 | 4.22E-02 |
| Seed length | 5.41E-06 | 8.07E-06 | 4.31E-03 | 1.84E-02 |
| Seed width | 1.69E-04 | 3.81E-04 | 1.93E-03 | 3.17E-02 |
| Seed width/length ratio | 7.17E-05 | 1.96E-04 | 4.29E-03 | 2.89E-02 |
| Seed coat luster | 1.78E-04 | 1.64E-04 | 3.50E-03 | 2.16E-02 |
| Flower color | 3.92E-06 | 8.38E-05 | 1.40E-03 | 1.79E-02 |
| Cotyledon color | 2.87E-05 | 1.68E-05 | 1.16E-05 | 4.51E-03 |
| Stem color | 5.96E-06 | 8.52E-05 | 1.48E-03 | 4.53E-03 |
| Pod color | 3.88E-06 | 1.27E-04 | 9.05E-04 | 4.35E-03 |
| Downy mildew resistance | 1.39E-04 | 1.88E-04 | 4.08E-05 | 5.21E-03 |
| Anthracnose resistance | 1.56E-05 | 1.53E-05 | 1.50E-03 | 4.22E-03 |
| Rust resistance | 2.62E-05 | 1.56E-05 | 3.21E-03 | 3.77E-03 |

¹Trait analyzed for GUA_2015. ²Mixed model control both relatedness and population structure.

³EMMA control for relatedness. ⁴6PC control population structure based on 6 principal components (25% of variation). ⁵Lowest MSD for each trait are in bold.

Genomics regions associated with altitudinal adaptation in population GUA_2015

The altitude where the genotypes from GUA_2015 were grown ranged from 1183 to 3050 masl, with an average of 2000 masl. Genomic regions controlling altitudinal adaptation were found in Pv02, Pv03, Pv05 and Pv11 (Figure 8). Pv02 has two significant peaks, the first one encompasses an interval of 76 Kb and the second peak encompasses a region of 121 Kb. The two peaks are 6Mb one from another, and for that reason, they were evaluated separately. In Pv05, peaks were also evaluated separate because there was a region of 7Mb between them. The distance encompassed by significant SNPs in Pv11 was 317 Kb. Candidate genes for altitudinal adaptation are listed in Table A2 after a search of 50 kb downstream and 50 kb upstream of the significant SNP in the V2. *Phaseolus v.* genome assembly and annotation. Variation explained by significant markers (R^2) was calculated for each significant peak found. Peaks in Pv02 explained 12% and 9% of phenotypic variation, respectively. Peak in Pv03 explained 8% of the variation, while each peak in Pv05 explained 7% of the variation. Finally, a significant peak in Pv11 explained 9% of the phenotypic variation.

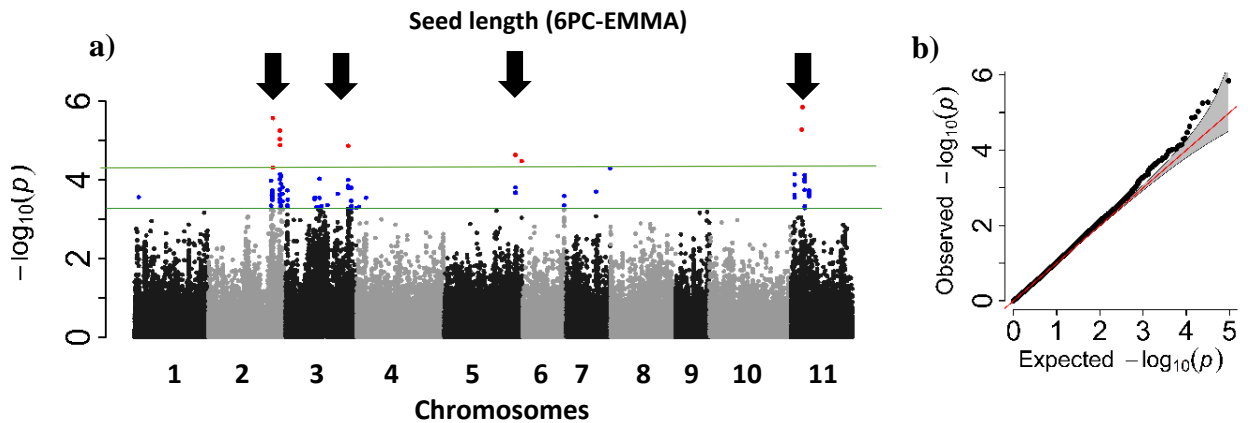


Figure 8. a) Manhattan plot for altitudinal adaptation of GUA_2015. b) quantile-quantile plot for the best model. Green lines represent the cutoff values for 0.1 and 0.01 percentiles. Markers significant for the 0.01 and 0.1 percentiles are colored in red and blue, respectively. Best model is indicated in parenthesis.

Genomics regions associated with seed length, seed width and seed width/length ratio in population GUA_1966-82.

Length of the seeds measured ranged from 8.6 to 14.9 mm, with an average of 11.4 ± 0.55 mm. Width of the seeds ranged from 3.6 to 7.7 mm, with an average of 5.0 ± 0.35 mm. The width/length ratio ranged from 0.3 to 0.8 with an average of 0.45 ± 0.03 . An important genomic region was found at the end of Pv02 associated with seed length, seed width and the seed width/length ratio (Figure 9). For seed length, a minor peak was found in Pv04. Table A2 shows the significant markers with the candidate gene models that best explained the traits. In the case of seed length, the significant peak in Pv02 explained 11% of the phenotypic variation while the significant peak in Pv04 explained 5% of the variation. For seed width, the peak in Pv02 explained 11% of phenotypic variation and in the seed width/length ratio the significant peak in Pv02 explained 25% of variation. The strongest signal for seed length in Pv02 covered a region of 535 kb, while for seed width the signal on Pv02 covered a region of 282 kb. However, the signal for both traits was located in the same region. As expected, the same region was identified for the seed width/length ratio.

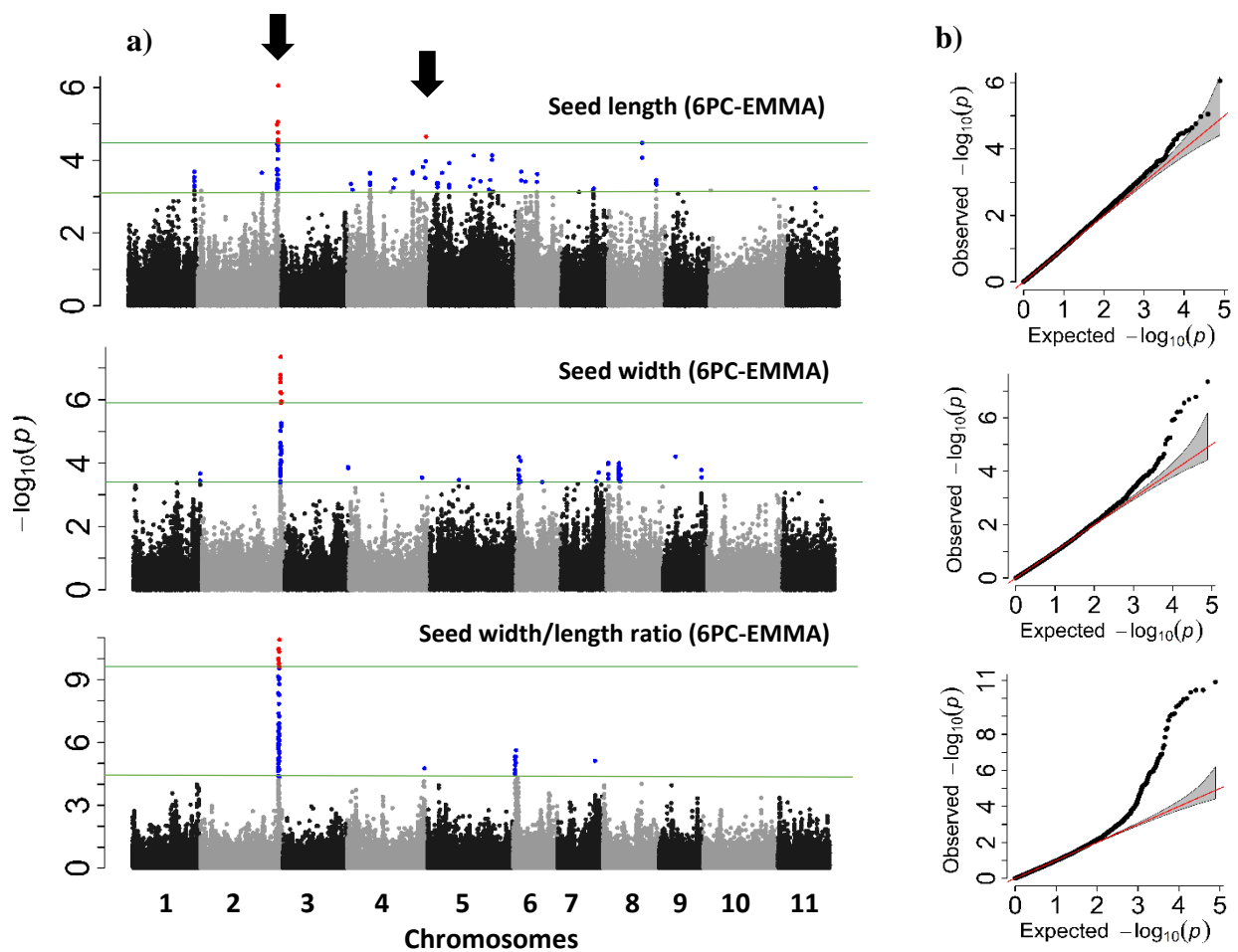


Figure 9. a) Manhattan plots for seed length, seed width and seed width/length ratio in accessions from GUA_1966-82. b) quantile-quantile plots for the best models. Green lines represent the cutoff values for 0.1 and 0.01 percentiles. Markers significant for the 0.01 and 0.1 percentiles are colored in red and blue, respectively. Best model is indicated in parenthesis.

Genomic regions associated with flower color, cotyledon color and stem color

Phenotypic results show that the Guatemalan climbing bean collection has mostly purple flowers but white, lilac and pink flowers are also found. In the case of cotyledon color, yellow is the predominant color, but also green, purple and brown are found. For stem, the colors found are green, green with pink stripes and green with purple stripes.

Genomic regions associated with flower color, cotyledon color and stem color were found at the beginning of Pv06 (Figure 10). For cotyledon color, additional genomic regions

were found in Pv02, Pv03, Pv04, Pv05, and Pv08. The candidate genes for significant peaks of these traits are listed in Table A2. In Pv06, the significant peak present in the three traits encompass a region of 5Kb. A second peak in the same chromosome is shared between flower color and stem color, and embrace an interval of 13Kb. However, no candidate gene was found around this region. Additional peaks were found in Pv06 for flower color and stem color. However, none of these peaks were close to gene models related to flower or stem color. Peaks in Pv04 and Pv08 for cotyledon color were also distant from any candidate gene.

The significant peak shared by the three traits explained 9%, 6% and 6% of phenotypic variation for flower color, cotyledon color and stem color, respectively. It seems that for flower color, both peaks (13Kb and 5Kb) explain the same phenotypic variation, and this may explain why no candidate gene was found for the second peak. The significant peak in Pv02 for cotyledon color explains 5% of the phenotypic variation.

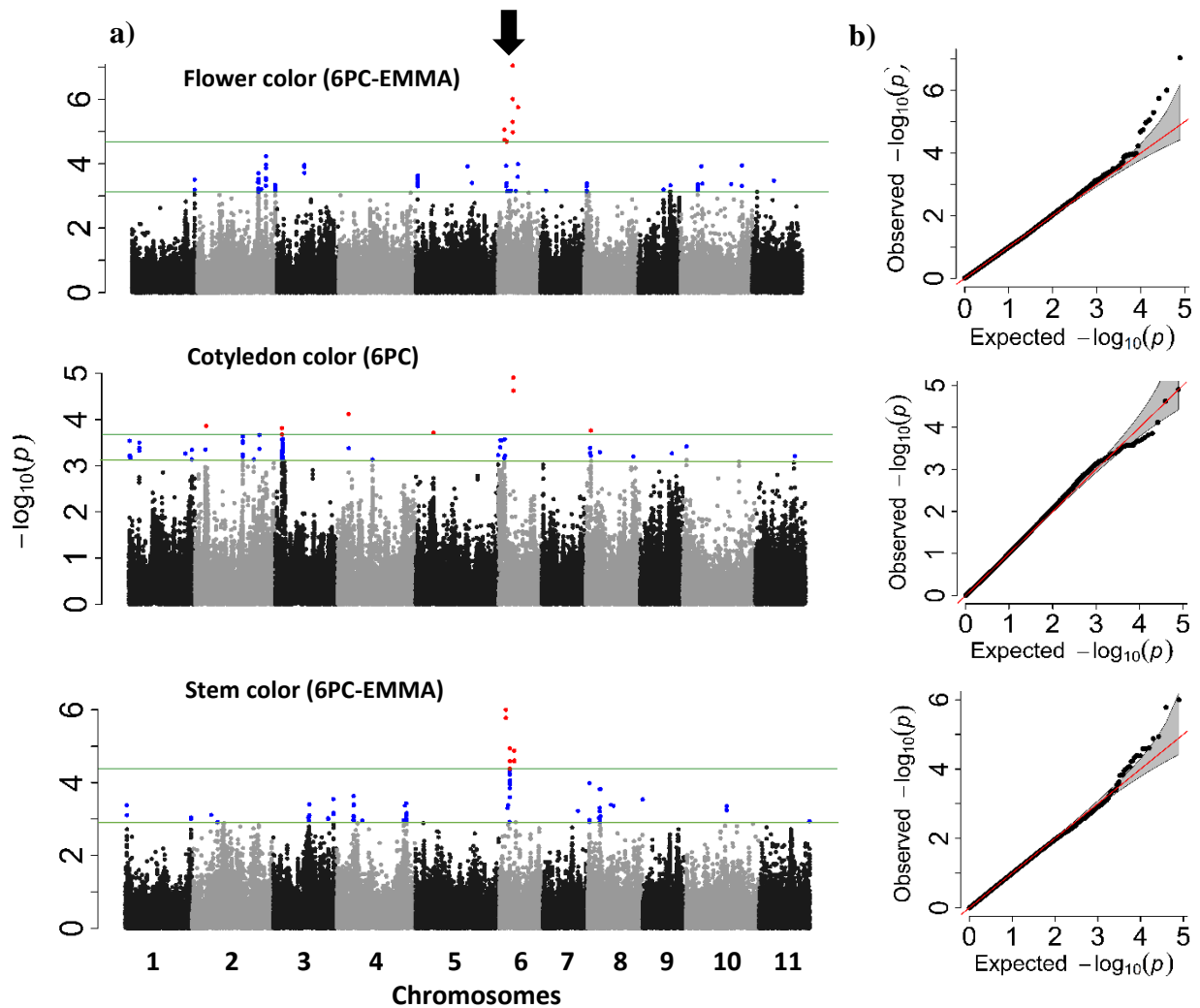


Figure 10. a) Manhattan plots for flower color, cotyledon color and stem color in accessions of GUA_1966-82. b) quantile-quantile plots for the best model. Green lines represent the cutoff values for 0.1 and 0.01 percentiles. Markers significant for the 0.01 and 0.1 percentiles are colored in red and blue, respectively. Best model is indicated in parenthesis.

Genomic regions associated with seed coat luster in population GUA_1966-82

Phenotypic results show that almost 79% of the Guatemalan climbing bean seeds are shiny and the rest are dull seeds. Two significant regions were found in Pv02 and Pv07 associated with seed coat luster (Figure 11). The peak in Pv02 spans an interval of 282Kb while the peak in Pv07 covers a region of 20Kb. Phenotypic variation explained by the peak in Pv02 is

6%, and the phenotypic variation explained by the peak in Pv07 is 8%. Candidate genes for seed coat luster are listed on Table A2.

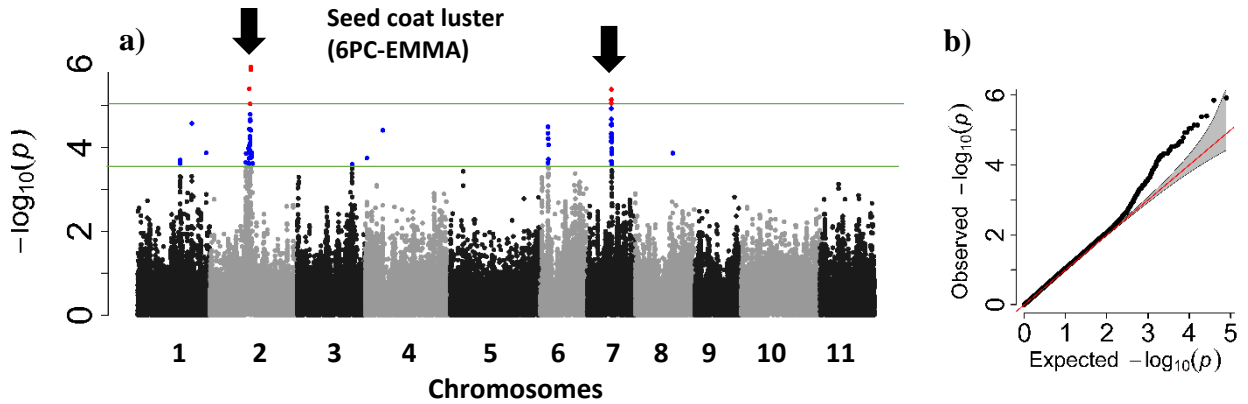


Figure 11. a) Manhattan plot for seed coat appearance in accessions of GUA_1966-82. b) quantile-quantile plot for the best model. Green lines represent the cutoff values for 0.1 and 0.01 percentiles. Markers significant for the 0.01 and 0.1 percentiles are colored in red and blue, respectively. Best model is indicated in parenthesis

Genomic regions associated with pod color in population GUA_1966-82

Pod color in the Guatemalan collection is mostly green. However, yellow, yellow with brown stripes, yellow with purple stripes, and purple pods are also found. A significant genomic region associated with pod color was found at the beginning of Pv08 (Figure 12). The region encompass an interval of 122bp. The candidate gene for this peak is described in table A.2. Total phenotypic variation explained by the markers in the significant peak is 6%.

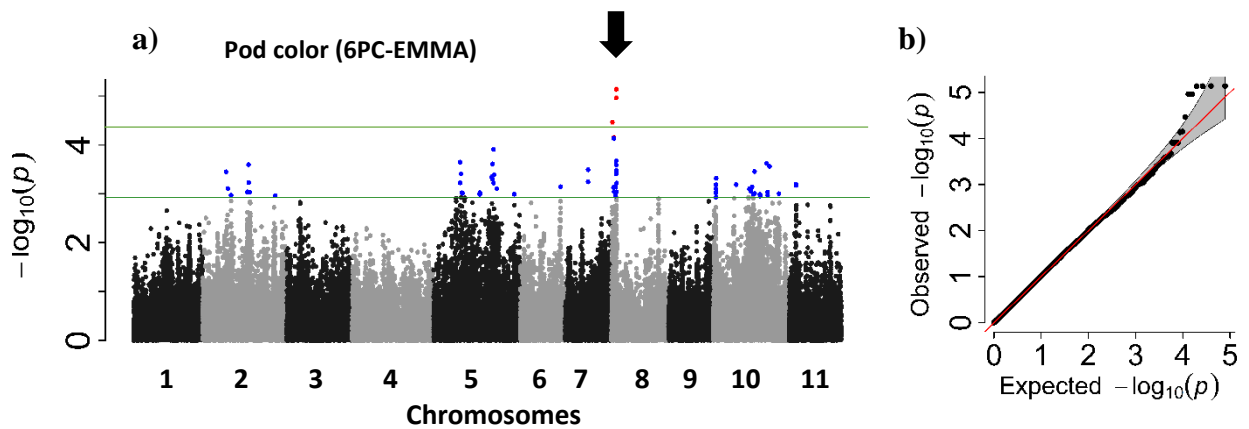


Figure 12. a) Manhattan plot for pod color in accessions of GUA_1966-82. b) quantile-quantile plot for the best model. Green lines represent the cutoff values for 0.1 and 0.01 percentiles. Markers significant for the 0.01 and 0.1 percentiles are colored in red and blue, respectively. Best model is in indicated parenthesis.

Genomic regions associated with rust (*Uromyces appendiculatus* (Pers) Unger) resistance in population GUA_1966-82

Phenotypic data for rust resistance show that only 18% of the population show complete resistance, and it has a normal distribution. Genomic regions associated with resistance to rust were found in Pv02 and Pv08 (Figure 13). Markers in Pv02 are located in a region of 332 Kb, while markers found in Pv08 are located 25 Kb one from another. Table A2 shows the candidate genes proposed for this trait. Phenotypic variation explained by the significant peak in Pv02 is 6%, while variation explained by the significant peak in Pv08 is 4%. These results suggest that resistance to rust is controlled by many factors.

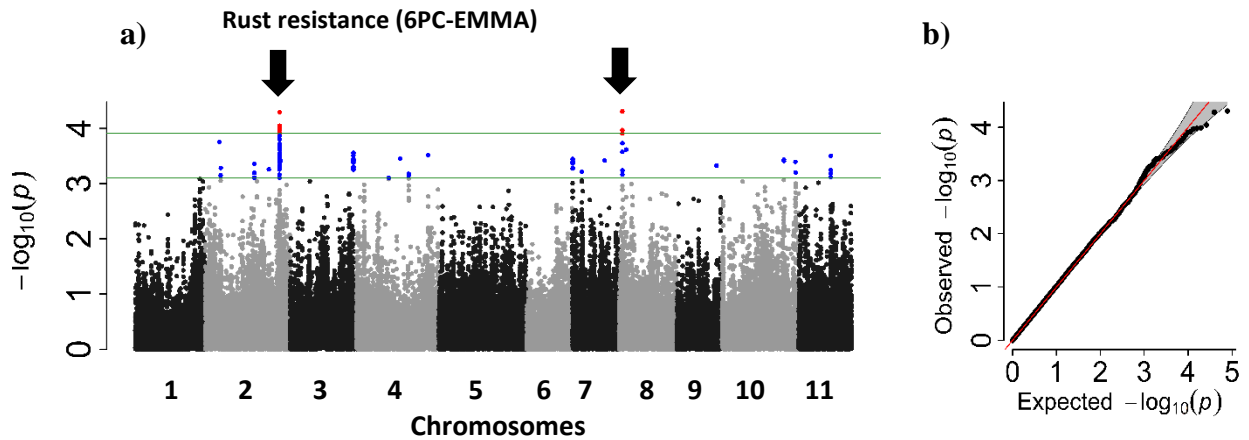


Figure 13. a) Manhattan plot for rust resistance in accessions of GUA_1966-82. b) quantile-quantile plot for the best model. Green lines represent the cutoff values for 0.1 and 0.01 percentiles. Markers significant for the 0.01 and 0.1 percentiles are colored in red and blue, respectively. Best model is indicated in parenthesis.

Genomic regions associated with anthracnose (*Colletotrichum lindemuthianum*) resistance in population GUA_1966-82

Contrary to rust, 50% of the Guatemalan accessions show resistance to anthracnose, and just a few genotypes are completely susceptible. Two genomic regions associated with anthracnose resistance were found in Pv01. Minor peaks were found in Pv03 and Pv05 (Figure 14). Significant markers in Pv01 span a region of 2.5 Mb. Phenotypic variation explained by the first peak in Pv01 is 9%. The second peak in Pv01 explained 5% of the variation, and peaks in Pv03 and Pv05 explained 5% of the variation each one. Candidate genes in these genomic regions are listed in table A2.

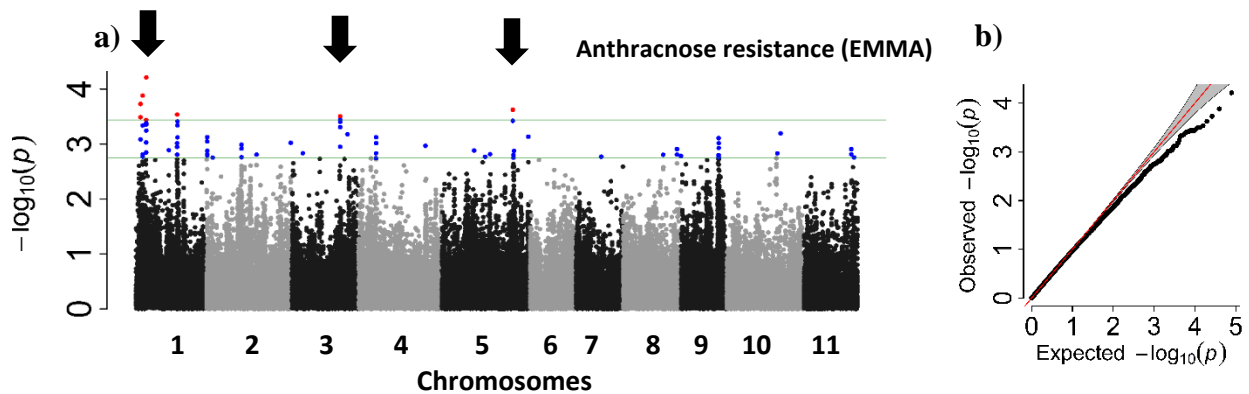


Figure 14. a) Manhattan plot for anthracnose resistance in accessions of GUA_1966-82. b) quantile-quantile plot for the best model. Green lines represent the cutoff values for 0.1 and 0.01 percentiles. Markers significant for the 0.01 and 0.1 percentiles are colored in red and blue, respectively. Best model is in parenthesis.

Genomic regions associated with downy mildew (*Phytophthora nicotianae*) resistance in population GUA_1966-82

Phenotypic data shows that 80% of the Guatemalan population is completely susceptible to downy mildew, and just 5 genotypes are resistant. Genomics regions associated with downy mildew resistance were located in Pv04 and Pv06 (Figure 15). Significant markers in Pv06 are located in a region of 94 Kb, while in Pv04 significant markers are located very close one from the other (3 bp). Phenotypic variation explained by the significant peak in Pv04 is 5% and for Pv06 is 13%. Candidate genes proposed for this trait are listed in table A2.

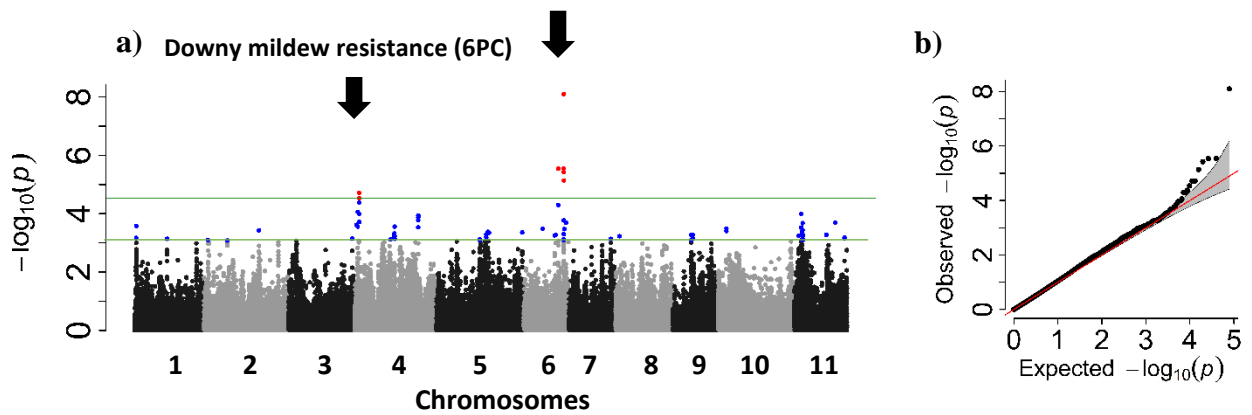


Figure 15. a) Manhattan plot for downy mildew resistance in accessions of GUA_1966-82. b) quantile-quantile plot for the best model. Green lines represent the cutoff values for 0.1 and 0.01 percentiles. Markers significant for the 0.01 and 0.1 percentiles are colored in red and blue, respectively. Best model is indicated in parenthesis.

DISCUSSION

Genetic Diversity and Population Structure

The analysis of two large collections of Guatemalan climbing beans has allowed us to confirm the unique diversity of this group. All evidence generated by these analyses demonstrates that Guatemalan climbing beans can be classified as another race within the Middle American gene pool.

Gene pools differentiation and race structure as proposed by Singh et al. (1991a), Beebe et al. (2000) and Blair et al. (2009) was confirmed in this study. We found that the Middle American gene pool and the Andean gene pool are highly differentiated. Race structure in the Middle American gene pool includes races Mesoamerica, Durango-Jalisco and Guatemala. Our study also found greater admixture between races of the Middle American gene pool and very low admixture between gene pools. Geographic proximity between races and breeding within market classes can explain this admixture. Misclassification of accessions found in this study can be explained by possible germplasm introgression from different regions.

The Middle American gene pool has been described as more diverse than the Andean gene pool (Chacón et al., 2005; Rossi et al., 2009; Bitocchi et al., 2012; Mamidi et al., 2013; Blair et al., 2013). In this study we found greater diversity for Middle American gene pool races than the Andean gene pool population that was evaluated. However, it is necessary to consider that the ADP population is mostly represented by accessions from race Nueva Granada, and diversity from races Peru and Chile could have been underrepresented. Additionally, we used fewer accessions to represent the Andean gene pool than the Middle American gene pool.

When we compare our diversity results with similar diversity studies using SNPs (Blair et al., 2013; Rodríguez et al., 2016), it can be observed that our study was able to identify greater

diversity in both the Middle American and Andean gene pools. However, all studies consistently observed a higher diversity in the Middle American gene pool. Differences in diversity can be explained by the number of SNPs used for each study. This would be in agreement with Müller et al. (2015) who suggested that at least three times more SNPs are needed than SSRs to obtain similar estimates of diversity. In all the cases, when we compared our study with diversity studies that used SSRs, diversity was lower using SNPs.

Independent domestication events for the Middle American and Andean gene pools have been widely discussed. Chloroplast haplotypes of races Mesoamerica, Durango, Jalisco and Guatemala are closely related and distantly related to chloroplast haplotypes of Andean races. Organelle DNA evidence supports a single domestication event in each gene pool (Chacón et al., 2005). In their diversity study, Kwak and Gepts (2009) found that the Middle American races cluster together, suggesting a single domestication for this gene pool and a later development of races. Similar results were found by Papa and Gepts (2003) who suggested that races were formed by farmer selection and ecological and geographical features rather than by many domestications events in the Middle American gene pool. McClean et al. (2012) suggested that difference in F_{ST} values between markers located close to domestication loci and markers located far from domestications loci result after domestication events, supporting a single domestication event in each gene pool. We found that population differentiation was the highest when gene pools were compared, and moderate when comparing between Middle American gene pool races, including race Guatemala. These results confirm the existence of race Guatemala in the Middle American gene pool, and also that greater diversity is found in this gene pool. Race Guatemala cluster with the other Middle American races and separate from the Andean gene

pool, suggesting that it appeared after the domestication event in the Middle American gene pool similar to the other races.

The percentage of heterozygous genotypes was high (31%) for the entire population (Andean and Middle American gene pools). This suggests that SNPs can detect higher amount of heterozygotes compared to other molecular markers (Blair et al., 2013). However, other studies with SNPs have reported a low percentage of heterozygotes (<0.01) for common bean (Müller et al., 2015). This discrepancy may suggest that landraces of Guatemalan climbing beans have a high amount of heterozygous loci despite their self-pollinated nature. The percentage of outcrossing in common bean can range from 1 to 70% depending on the conditions where it grows (Rendón-Anaya et al., 2017), and if it was high in the Guatemalan highlands a higher amount of heterozygotes would be observed. However, a high percentage of heterozygotes was observed across all the subpopulations evaluated here, especially within race DJ. This observation agrees with the STRUCTURE analysis results where a higher amount of race DJ accessions were admixed compared with the admixed accessions found in race M.

When both Guatemalan collections are compared, we can observe that GUA_2015 is more diverse and admixed than GUA_1966-82. However, pairwise F_{ST} values and the AMOVA show that both subpopulations are very similar. Unfortunately, passport data is not available for GUA_1966-82, which seems to be a very unique germplasm population. Beebe et al. (2000) described race Guatemala as the least diverse. However, they used a small population size of climbing beans for this study. On the other hand, Blair et al. (2009) described race Guatemala as one of the most diverse races in the Middle American gene pool. Our results for GUA_1966-82 agree with Beebe et al. (2000), while GUA_2015 was the most diverse population.

Results of STRUCTURE analysis showed that GUA_2015 has more admixture than GUA_1966-82. GUA_2015 also has more introgression of DJ than M, even though race M is predominant for the bush type beans in Guatemala. There is very little information about climbing beans from Guatemala. De Young et al. (2017) collected data in a survey performed in the highlands of the country, where smallholders stated that only 6% of them use an improved variety and they also reported that they have used the same market class and recycled seed for over 23 years. It seems that introgression of DJ in the Guatemalan climbing beans is recent, since the old collection does not show this introgression. The nature of this introgression is unknown especially since this region has not been targeted for the development and release of new varieties specifically adapted to this region. Interestingly, Blair et al (2009), described in their analysis that race Guatemala and race DJ had the lower genetic differentiation between them and suggested that these races may have shared ancestry.

The GUA_2015 collection is also unique because it shows introgression of material that is not present in any of the common bean races. Since this common bean race is grown in close proximity with genotypes from the domesticated *P. dumosus* and *P. coccineus* species, natural crosses between these and *P. vulgaris* may have generated the introgression observed. However, a SNP diversity analysis of these other species need to be performed.

It was observed that GUA_2015 was 3% more diverse when compared to the wild accessions. In their study, Papa and Gepts (2003) reported that wild genotypes from Central America (Guatemala and Costa Rica) were the least diverse among all the wild accessions from the Middle American gene pool. Since we use only wild accessions from Guatemala, this may explain why we found genetic gain in the GUA_2015 group. However, introgression between

this newer population and wild germplasm may provide an explanation for the greater diversity for this newer population.

Intra-accession diversity analysis of the Guatemalan beans showed that there is a low percentage of heterozygotes within accessions, but moderate genetic diversity is still found. Despite variation among accessions is high, within accessions is also relatively high. Species *P. dumosus* and *P. coccineus* have a major percentage of outcrossing compared to *P. vulgaris*. Since these species were included in the intra-accession analysis, it can partially explain why we found moderate genetic diversity. Another possible reason that explains the amount of diversity is that genotypes analyzed are mostly landraces, since breeding efforts to improve Guatemalan climbing beans was only initiated a few years ago. Just recently, and for the first time, two improved varieties of climbing beans (ICTA Utatlan and ICTA Labor Ovalle) were released by ICTA in collaboration with NDSU at the beginning of 2017, and their features included less aggressiveness when grown in combination with maize (LIL, 2017).

Genome Wide Association Study

Altitudinal adaptation

Selection, gene flow and genetic drift are major factors leading to local adaptation. Climatic and geographic factors, and biotic and abiotic stresses are factors that drive these three effects and lead to local adaptation and fitness of the species (Joshi et al., 2001). In Guatemala, altitude is an important geographic factor influencing local/regional climate. In the country, altitude can range from 0 to more than 3,000 masl in an area of 108,889 km² (Lachniet and Patterson, 2009). Moreover, in our analysis of the GUA_2015 population, we found that altitude is an important factor that can explain population structure. Flowering time and leaf length are

two traits that appear to exhibit phenotypic differentiation because of local adaptation (Klaedtke et al., 2017; Romero-Navarro et al., 2017).

In this study, genomic regions associated with altitudinal adaptation in the Guatemalan climbing beans were identified using elevation data as the dependent variable. The significant SNP Pv02_38216242 fell inside the candidate gene Phvul.002G213600, which encodes a RNA-binding KH domain-containing protein (HEN4). The significant SNP Pv02_38337312 is also close to the gene model Phvul.002G215500 that encodes an AGAMOUS-like 21 protein (AGL). AGL is a gene that participates in the differentiation of stamen and carpel cells. The presence of AGL in the floral meristem is regulated by positive and negative RNA transcription regulators (Lenhard et al., 2001; Franks et al., 2002). However, it has been suggested that AGL can also be controlled at a posttranscriptional level (Cheng et al., 2003). HEN4, HUA1 and HUA2 are proteins involved in the differentiation of floral reproductive organs via processing of AGL mRNA, and HEN4 is directly linked to differentiation of stamen cells and conditions floral determinacy (Cheng et al., 2003).

Flowering time is regulated by different factors including photoperiod, circadian clock and vernalization. Different protein pathways regulate the response of the plant to these factors (Blumel et al., 2015). One of the significant markers found on Pv11 (Pv11_13054356) is 52kb from the gene model Phvul.011G105600 that encodes a homeodomain-like superfamily protein, also known as LATE ELONGATED HYPOCOTYL (LHY). LHY is a gene involved in the photoperiodic induction of flowering in *Arabidopsis* (Mouradov et al., 2002). The CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) and the LHY genes regulate the expression of the TIMING OF CAB EXPRESSION 1 (TOC1) gene, which is related to the photoperiod response (Alabadi et al., 2001). This interaction generates a negative regulation of the photoperiod flowering

pathway (Niwa et al., 2007). This suggests that the expression of LHY affects flowering time based on the duration of the light period during the day because overexpression of the gene generates late flowering during long day plants (Mouradov et al., 2002). In general, AGL, HEN4 and LHY have important roles in flowering time, suggesting that flowering time is also an important trait for the altitudinal adaptation of the Guatemalan climbing beans.

Significant SNP Pv02_46664311 was found inside the gene model Phvul.002G298300 that encodes a RING/U-box superfamily protein. The ubiquitin/26S proteasome pathway is involved in most of the cellular processes, including circadian rhythms, hormone signaling, and biotic and abiotic stress responses (Vierstra, 2003). The ubiquitin protein is interceded by three enzymes (Ubiquitin-activating enzyme, ubiquitin-conjugating enzyme and ubiquitin protein ligase). The third enzyme can be of the RING, F-box or U-box type (Yee and Goring, 2009). U-box proteins have been related specifically with stress responses that are up-regulated in the presence of cold, drought, osmotic stress and other abiotic stresses, factors that are related to adaptation to ecological pressures (Yee and Goring, 2009). Hypersensitive response (HR) has also been related to the upregulation of U-box genes. In parsley (*Petroselinum crispum*. (Mill.) Fuss), the gene activated in the first level of defense response was PcCMPG1 (a U-box gene) (Kirsch et al., 2001).

Two significant markers in Pv02 (Pv02_46736576 and Pv02_46740440) fell 7kb close to the candidate gene Phvul.002G299400 that encodes a calcineurin B-like protein 8 (CBL8). CBL proteins, similar to U-box proteins, are related to stress responses but via interaction with CBL-interacting protein kinases (CIPKs) (Pandey et al., 2015). Interestingly, CBL19 in Arabidopsis has also been related to pollen tube growth and polarity, which suggests it also participates in the flowering process (Zhou et al., 2015). Pandey et al., (2015) determined that CBL2 and CBL3,

while interacting with CIPK21, control the responses to salt stress conditions. Similarly, upregulation of CIPK31 in the grass species *Brachypodium distachyon* (L.) P. Beauv. improved the tolerance to salt and drought, via stomatal closure (Luo et al., 2017).

In Pv03, one significant SNP (Pv03_39752668) was located 35kb from the gene model Phvul.003G175700, which encodes a dicer-like 4 protein (DCL4). The function of the DCL genes, is the biogenesis of small RNAs that negatively interfere with the RNA process in a way to regulate gene expression (Xie et al., 2005). Small RNAs are classified in micro-RNAs and small interfering RNAs (siRNAs). In their study, Xie et al. (2005) proposed that DCL4 regulates the trans-acting siRNAs, which accelerates the vegetative phase change and shortens the life cycle of the plant. The change of vegetative phases is also related to the start of the reproductive phase in plants (flowering time). DCL4 has also been suggested to have antiviral response (Deleris et al., 2006). As mentioned above, local adaptation of species is related to the response of biotic stresses including viruses that can be present in specific regions.

One significant SNP located in Pv05 (Pv05_32602302) was found 3kb near to the candidate gene Phvul005G106100 that encodes a protein phosphatase 2C family protein (PP2C). One of the most important pathways for biotic and abiotic stresses responses, which ensures the rapid adaptation to different environments in plants, is the abscisic acid (ABA) signaling pathway (Sirichandra et al., 2009). Genes related to the ABA pathway are upregulated when the plant is submitted to osmotic stresses (Nambara and Marion-Poll, 2005). The PP2C family proteins are known as regulators of the ABA pathway (Umezawa et al., 2009) and therefore related with the plant response to biotic and abiotic stresses.

The second most significant marker found on Pv05 (Pv05_39204798) is located 3kb near the gene model Phvul.005G164300 that encodes a leucine-rich repeat protein kinase family

protein (LRR-PK). Genes associated with defense response in constantly evolve to overcome susceptibility to pathogens (Brown, 2003). Pathogen-associated-molecular-pattern (PAMP)-triggered immunity (PTI) is a nonspecific pathogen defense response and can be general to many pathogens. It triggers immunity by recognizing PAMPs like flagellin in bacteria or chitin in fungi (Nicaise et al., 2009). PTI is considered less aggressive than effector-trigger-immunity (ETI), where pathogen specific effectors are recognized by the plant, which then activate the HR (Rafiqi et al., 2009). LRR-PK genes are usually associated with PTI. However, it is believed that there is an important link between PTI and ETI since proteins involved in PTI are targets of the pathogen effectors that activate ETI (Chisholm et al., 2006). Interestingly, in common bean, a virulent strain of anthracnose decreased when different altitudinal distances were evaluated (Sicard et al., 2007). These results support the possibility of local adaptation of common bean for resistance to most common strains of pathogens found in the region. In summary, significant genetic factors associated with altitudinal adaptation in Guatemalan climbing beans map near genes related to flowering time and response to biotic and abiotic stresses.

Seed shape

Seed shape is indirectly related to seed yield and seed weight. Most of Guatemalan climbing beans have a round shape. However, the GUA_1966-82 populations segregated for both round and flat seed shape. This provided an ideal population to map genes associated with seed shape. Markers associated with seed size traits have been reported in all chromosomes except Pv08 and Pv10 (Blair et al., 2009). In this study, SNPs related to seed shape and seed size colocalized signals at the end of Pv02. The signal at Pv02, clusters a group of SNPs that fell inside different candidate genes that may explain differences in seed shape. Pv02_47556089 is a SNP within the gene model Phvul.002G307600, which encodes for a profilin protein (PRF).

Profilin is a small actin-binding protein. Actin forms the cytoskeleton of plant cells and have important functions in many cellular processes, including cell expansion (Baskin 2005). In Arabidopsis, PRF1 is involved in cell shape and cell elongation (Ramachandran et al., 2000). Profilin is also necessary for tip growth in the moss *Physcomitrella patens* (Hedw.) Bruch & Schimp. (Vidali et al., 2007). Cao et al. (2016) found that PRF1 coordinates the organization of actin filaments during cell expansion. No report about the effect of profilins in seed size and shape exists. However, as a protein involved in cell shape it has a direct effect in tissue shape.

SNP Pv02_47021193 fell inside a gene model that encodes a SH3 domain-containing protein (Phvul.002G302000). It has been suggested that SH3 domains interact with cytoskeletal molecules and GTP binding proteins (Matuoka et al., 1993). In their study, the actin accumulation generated a change in human cells shape. In plants, the function of SH3 domain-containing proteins has been demonstrated to be similar to one in yeast that is involved in the arrangement of actin and vesicle trafficking via cytoskeleton (Lam et al., 2001).

SNP Pv02_47294046 is located 6kb from Phvul.002G304600, which encodes a pyridoxal phosphate phosphatase-related protein. A similar SNP is found for seed width, where Pv02_47516312 is found near Phvul.002G307100, which encodes a phosphotyrosine protein phosphatase superfamily protein. Protein phosphatase regulatory pathways have been reported to have important effects on grain length and grain width in *Oryza sativa* (Zheng et al., 2015). *qGL3/GL3.1* is a rice QTL that encodes a serine/threonine phosphatase related with grain length and grain weight (Qi et al., 2012; Zhang et al., 2012).

Two SNPs, Pv02_47352798 and Pv02_47352799 were found within the gene model Phvul.002G305400, which encodes an F-box/RNI-like superfamily protein. This protein, known as SKP2A protein in Arabidopsis, has been suggested to regulate a transcription factor that

regulates cell multiplication. The ubiquitin/26S proteasome pathway is related to grain shape in rice (Song et al., 2007) and to seed size of plants in general by negatively regulating cell division (Du et al., 2014). As mentioned above, the ubiquitin protein is interceded by three enzymes (ubiquitin-activating enzyme, ubiquitin-conjugating enzyme and ubiquitin protein ligase). The third enzyme can be of the type SKP or F-box proteins (Smalle and Vierstra, 2004). In their study, Jurado et al. (2008) reported that SKP2A regulates a transcription factor that inhibits the cell division, and therefore the overexpression of SKP2A increases the number of cells.

Three SNPs for seed width (Pv02_47234713, Pv02_47234716 and Pv02_47234824) fell inside the gene model Phvul.002G304100, which encodes a tetraspanin protein. Similar to the ubiquitin/26S pathway, TET5 and TET6 have been suggested to have a negative role on cell proliferation in roots and leaves (Wang et al., 2015). Tetraspanins take part in important processes of the cell like morphogenesis (Yáñez-Mó et al., 2009), cell specification and some molecular pathways (Wang et al., 2015).

Two markers for seed width are found close to a protein kinase with adenine nucleotide alpha hydrolases-like domain (Pv02_47262606, Pv02_47273758). OsAGSW1 is a protein kinase that has been ligated to grain size and weight in rice. Overexpression of this protein leads to an increase on the size of the grain and therefore an increase in grain weight, via regulation of cell division and generation of vascular bundles (Li et al., 2015).

A single significant marker associated with seed width (Pv02_47400959) could not be linked with any candidate gene because no genes were found in the 100kb interval determined for the search of gene models. In general, genes found to be associated with seed shape in other species are usually regulators or modifiers of other proteins. In our study, candidate genes for seed shape are associated with cell size, cell expansion or cell proliferation.

Flower color, cotyledon color and stem color

In common bean, a large set of genes have been described that control color in seed coats, flowers, stems, and pods. Most of these genes are related to the biosynthesis of flavonols and anthocyanins. Final colors depend on the level of anthocyanins accumulated in the tissue (McClellan et al., 2002). Within these genes, the *P* gene is the most important because it defines colored or not colored (white) seed coats and flowers (Bassett 1996). Other genes described for color include *C*, *R*, *J*, *D*, *G*, *B*, *V* and *Rk* genes (Bassett et al., 1990; McClellan et al., 2002; Yuste-Lisbona et al., 2014). The *V* gene, in presence of a dominant *P* gene, determines the purple and pink color in flowers (Beninger et al., 1999). The *V* gene also interacts with the other genes to generate a broad range of seed colors known in common bean (Prakken, 1972). It has also been suggested that this gene can control pod, stem and leaf color (Bassett, 2005). The *V* gene was previously identified and located at the beginning of Pv06 (McClellan et al., 2002). 21 cosegregating indel markers for the *V* gene were identified in the pericentromeric region of Pv06, but no recombinants were found in this 14 Mb region (McClellan and Lee, unpublished data).

In our study, a significant peak was found at the beginning of Pv06 for flower color, cotyledon color, and stem color. Significant SNPs Pv06_7000786, Pv06_7001628, Pv06_7001680 and Pv06_7006012 were found 46kb near the candidate gene Phvul.006G15400, which encodes a cytochrome P450 superfamily protein (CYP75B). The anthocyanin biosynthesis pathway is divided into three stages: the phenylpropanoid pathway, the flavonoid pathway, and the anthocyanin pathway (Lepiniec et al., 2006; Olsen et al., 2008). The flavonoid 3'5'-hydroxylase (F3'5'H) is a protein that belongs to the CYP75B family and is related to the hydroxylation of the B-ring in the precursor of the anthocyanidins (Tanaka and Brugliera, 2013). The number of hydroxyl groups determines the intensity of the color in flowers and therefore the

F3'5'H determines the flower color. Plants lacking this protein are not able to produce purple flowers (Tanaka and Brugliera, 2013). We suggest that the signal found on Pv06 is associated with the *V* gene, and the CYP75B is a good candidate that needs to be studied in detail.

The significant marker found in Pv02 for cotyledon color (Pv02_9064264) is 20kb close to the gene model Phvul.002G065600, which encodes a SPFH/B and 7/PHB domain-containing membrane-associated protein family (HIR1). Jasmonic acid (JA) is a plant hormone that participates in many cell processes including biosynthesis of anthocyanins (Jeong et al., 2004). Recently, the hypersensitive induced reaction (HIR) protein was found to interact with the jasmonate ZIM-domain (JAZ) protein during anthocyanin biosynthesis (Chen et al., 2017). It is known that the JA pathway is related to the synthesis of anthocyanin because it upregulates the genes in charge of the synthesis of anthocyanins (Shan et al., 2009). In their study, Chen et al. (2017) found that the HIR protein interacts with the JAZ-HLH complex and acts as a negative regulator of the accumulation of anthocyanins in apple.

The other significant peaks in Pv06 for flower color and stem color were not linked to any candidate gene. It seems that all the peaks found for both traits explained the same phenotypic variation, with the exception of marker Pv06_11349339 that is 4Mb away from the other markers and explains 7% of the phenotypic variation. However, no candidate gene was linked to it. The low recombination region where the peak is located explains why no candidate genes can be found in an interval of 100kb. In the case of cotyledon color, peaks in Pv03 and Pv05 were 300Kb away from MYB domain protein 106 and a chalcone-flavanone isomerase family protein, respectively. Both genes have an important role in biosynthesis of anthocyanins (Qi et al., 2011a; Guo et al., 2015) but they were not analyzed because they are out of the physical range established for considering them as candidate genes.

Seed coat luster

Seed coat in common bean can be dull or shiny (Cichy et al., 2014). Over time, a few genes have been reported to control seed coat luster in common bean, including the *Asper* (*Asp*) and *Joker* (*J*) genes (Lamprecht, 1940; Basset, 1996). Bassett (1996) also found that the *Asp* gene controls the seed coat luster and is independent of the seed coat color genes. Seeds with a dull coat (*asp/asp*) have elevations in the epidermis palisade cells, the seed coat is thinner because the palisade cells are smaller, and they also have a lower amount of anthocyanins. In the other hand, when the *Asp* allele is present, the epidermis layer does not have elevations and the seed is shiny and smooth (Beninger et al., 2000). The *J* gene is not as common as the *Asp* gene but has been suggested that its recessive allele (*jj*) can reduce shininess even in presence of *Asp* (Konzen and Tsai, 2014). Seed coat luster has also been related to seed water uptake. Shiny seeds have a lower water absorption compared with dull seeds (Cichy et al., 2014; Konzen and Tsai, 2014). This is related to the thicker seed coat of the shiny seeds (Konzen and Tsai, 2014). Water absorption is important because processors prefer seeds with a quick water absorption, since it determines cooking time of the beans (Cichy et al., 2014). Interestingly, in an interview to 600 people in Guatemala, 18% had preference for shiny seeds (Diamant et al., 1989).

The *Asp* gene has been located in Pv07, while *J* has been located in Pv10 (Miklas et al., 2000; McClean et al., 2002). We found an important genomic region associated with seed luster on Pv07 (markers Pv07_17581536, Pv07_17581555, Pv07_17581604 and Pv07_17601477). Gene model Phvul007G119178, which encodes a bifunctional inhibitor/lipid-transfer protein/seed storage2S albumin superfamily protein was 88Kb away from these markers. This protein is related to lipid transfer. Shiny seeds have a thicker layer of wax, reducing their capacity for water absorption. It suggests that the transference of lipids may influence the width

of the wax layer in shiny seeds. Cichy et al., (2014) proposed a FAE1/Type III polyketide synthase-like protein as candidate for the *Asp* gene. However, the gene model is located 12Mb away from the significant region that we found in this study.

The second significant peak found in this study was located on Pv02 (markers Pv02_23321712, Pv02_23604335, and Pv02_23604337). No genes related to seed coat luster has been reported in Pv02 before. The *B* gene for seed coat color has been mapped in Pv02 (Kelly and Vallejo, 2004). However, there is no evidence that it is related to seed coat luster. Interestingly, the significant marker Pv02_23321712 was close to a chalcone–flavanone isomerase family protein. As explained above, the synthesis of anthocyanins is divided into three stages. The second stage, the flavonoid pathway, needs the chalcone isomerases for the formation of dihydroflavonols (Shi and Xie, 2014). Interestingly, Beninger et al. (2000) suggested that the *B* gene functions similarly to chalcone isomerase in the production of flavonoids. It is necessary to evaluate if a gene related to seed color can also influence the luster of the seed in the Guatemalan climbing beans.

Pod color

Pod color can vary from green to purple when pods are immature, and it is linked to the level of anthocyanins accumulated on the pod tissue (Yuste-Lisbona et al., 2014). The *P* gene, which controls flower color, also defines the purple color in pods (Bassett, 1996). The *Prp* locus in common bean was first described by Okonkwo and Clayberg (1984). It was also studied by Bassett (2005), and he suggested that this locus determines whether the purple color is solid or in stripes. The location of the *Prp* locus was first designated by Kelly and Vallejos (2004) in Pv08. Yuste-Lisbona et al., (2014) identified QTLs for pod shape and pod color and found significant QTLs for pod color in Pv02, Pv06, Pv07 and Pv08.

In our study, we found a significant peak for pod color at the beginning of Pv08. All significant markers (Pv08_3153417, Pv08_3153521, Pv08_3153528, and Pv08_3153534) were located near the candidate gene Phvul.008G038000, which encodes a MYB domain protein 16 (MYB16). In the plant model *Arabidopsis thaliana*, the biosynthesis of anthocyanin pathway has been widely studied. These studies have demonstrated that the pathway is controlled by a MYB domain transcription factor (Shi and Xie, 2014). The anthocyanin biosynthesis pathway is divided in three stages (Olsen et al., 2008; Lepiniec et al., 2006). MYB transcription factors are related to the flavonoid pathway (Stracke et al., 2010). In their study, Wang et al., (2016) established that when tissue of apple (*Malus pumila* Miller) was exposed to light and high temperature, expression of MYB16 was activated and the anthocyanin biosynthesis was repressed. We suggest that the significant peak found in Pv08 can correspond to the *Prp* locus, and we propose MYB16 as the candidate gene for controlling pod color in the Guatemalan climbing beans.

Rust resistance

Bean rust is caused by the pathogen *Uromyces appendiculatus* and is one of the most important common bean diseases worldwide. It especially affects tropical areas (Stavely and Pastor-Corrales, 1989). Rust resistance in bean is usually determined by a gene-by-gene interaction (Flor, 1955). Nowadays, 10 genes of resistance have been reported in common bean: *Ur-3*, *Ur-4*, *Ur-5*, *Ur-6*, *Ur-7*, *Ur-9*, *Ur-11*, *Ur-12*, *Ur-13* and *Ur-14* (Augustin et al., 1972; Ballantyne, 1978; Stavely, 1984, 1990; Grafton et al., 1985; Finke et al., 1986; Jung et al., 1998; Liebenberg and Pretorius, 2004; de Souza et al., 2011), where five belong to the Middle American gene pool and five belong to the Andean gene pool (Hurtado-Gonzales et al., 2017). *Ur-13* has been reported to be located in Pv08, while no gene for rust resistance was reported in

Pv02 until recently. Montejó-Domínguez et al. (2017) reported a significant genomic region in Pv02 associated to rust resistance (races 20-3 and 31-1) evaluated in Guatemalan climbing beans.

A significant SNP for rust resistance (Pv02_42933517) was located inside a candidate gene that encodes for a leucine-rich repeat transmembrane protein kinase (LRR-PK) family protein (Phvul.002G256600), while a second significant SNP (Pv02_42893682) was located close to the same gene (38Kb). Resistance in plants are usually divided into two levels. In the first level, pathogen recognizers (PRRs) are located in the surface of the plant tissues and once they recognize molecular patterns of the pathogens they trigger the defense response known as PTI (Nicaise et al., 2009). These pathogen recognizers are usually protein kinases. Many LRR-RKs have been reported to recognize pathogen molecules. EFR is a LRR-RK that recognizes the bacterial elongation factor Tu in *Arabidopsis* (Zipfel et al., 2009). In rice, Xa21 is the LRR-RK that confers resistance to *Xanthomonas* (Song et al., 1995). Thus, the LRR-PK found in this study can have an important role in the first level of defense against *U. appendiculatus*.

The significant marker Pv02_42990094 was located 5kb near the gene model Phvul.002G257800, which encodes a RPM-interacting protein 4 (RIN4) family protein. RPM1 is a well-known resistance (R) gene in *Arabidopsis* that is associated with the “guard-guardee” hypothesis of pathogen recognition, where RPM1 acts as the guard of RIN4, which is usually the target of many pathogen effectors (Qi et al., 2011b). RIN4 is a protein required for seedling growth, meristem function and fertility, but has been also suggested that in *Arabidopsis* it represses PTI to activate the second level of resistance in plants (gene-for-gene interactions) and therefore it enhances the resistance to *Pseudomonas syringae* (Brown, 2003).

The last significant marker found in Pv02 for rust resistance was 623bp from the gene model Phvul.002G256600, which encodes an ubiquitin-specific protease 3 protein. Ubiquitin-

specific proteases have been suggested to take part in plant response to different stresses, including pathogen invasions (Zhou et al., 2017). They participate in protein deubiquitination in the Ubiquitin/26S proteasome pathway. PRRs are suggested to be degraded by the Ubiquitin/26S pathway. In tomato, AtUBP12 and AtUBP13 participate in immunity responses against *P. syringae* as a regulator of HR (Ewan et al., 2011). The region reported by Montejo-Dominguez et al. (2017) is 4Mb apart from the signal found in this study.

Two significant markers were found in the Pv08 region (Pv08_2277889 and Pv08_2303436), they are located at the beginning of the chromosome while *Ur-13* is located at the end of the chromosome. Interestingly, a significant SNP reported by Montejo-Dominguez et al. (2017) for rust resistance (race 20-3) in GUA_1966-82 is just 87bp away from Pv8_2303436. Both markers flank a cluster of candidate genes that encode Malectin/receptor-like protein kinase family proteins (Phvul.008G028200 and Phvul.008G028600). Similar to LRR-RKs, malectin/receptor-like protein kinases (known as FERONIA in Arabidopsis) have been ligated to the plant immune signaling cascade. FERONIA also participates in the gametophyte interactions during the pollen reception (Kessler et al., 2010) and controls the invasion of powdery mildew in Arabidopsis (Hückelhoven et al., 2013). In a recent study, Stegmann et al. (2017) described that FERONIA regulates the complex formed between EFR and FLAGELLIN-SENSING 2 to initiate the signaling cascade to activate immunity against *P. syringae* in Arabidopsis. In general, no R genes are proposed as candidates in our study, probably because in the phenotypic evaluation, no specific races of the pathogen were used. However, interesting links with protein kinases and proteins that may interact with R proteins were found.

Anthracnose resistance

Anthracnose is caused by the pathogen *Colletotrichum lindemuthianum*, and it represents one of the most important diseases in common bean (Ferreira et al., 2013). Similar to rust, the defense response mostly follows the gene-for-gene interaction (Flor, 1955). Many genes have been reported in common bean conferring resistance to the pathogen, including *Co-1* in Pv01 (Cardenas et al., 1964); *Co-2* in Pv11 (Mastenbroek, 1960); *Co-3* in Pv04 (Bannerot et al., 1971); *Co-4* in Pv08 (Fouilloux, 1976); *Co-5* in Pv07 (Fouilloux, 1976); *Co-6* in Pv07 (Goncalves-Vidigal et al., 1997); *Co-9* in Pv04 (Geffroy et al., 1999); *Co-10* in Pv04 (Alzate-Marin et al., 2003); *Co-13* in Pv03 (Concalves-Vidigal et al., 2008); *Co-u* in Pv02 (Geffroy et al., 2008); *Co-y* in Pv04 (Geffroy et al., 1999); and *Co-z* in Pv04 (Geffroy et al., 1999). Many of these reported genes have also reported alleles.

In our study, two significant markers (Pv01_3734105 and Pv01_3734112) were found within a gene model that encodes a protein kinase superfamily protein (Phvul.001G044900). As discussed above for rust resistance, protein kinases are involved in the first level of the plant defense response against pathogens. They start the signaling cascade needed to activate immunity.

A third significant SNP located in Pv01 (Pv01_4836007) was found near the candidate gene Phvul.001G046400, which encodes for a glycosyl hydrolase family protein with a chitinase insertion domain. Most of chitinases belong to the glycosyl hydrolases (Hjort et al., 2014). Chitin is one of the most important components of fungi cell wall (Chisholm et al., 2006). It is suggested that plants have evolved to recognize chitin as part of PTI to trigger defense responses. Even more interesting, plants do not produce chitin but have evolved to produce chitinases as a defense mechanism against fungi, to degrade the pathogens cell wall (Chisholm et al., 2006).

The significant SNP found in Pv03 (Pv03_33052812) was not near to any important candidate gene in the interval that was established. However, 58 Kb from the marker, gene model Phvul.003G134100 that encodes a MAP kinase 2 was found. Interestingly, Maldonado-Mota et al. (2017) found a significant region in Pv03 associated to anthracnose resistance for race 73 of the pathogen, this region includes the marker found in this study (Pv03_33052812). MAP kinases are involved in the signaling pathway that activates resistance against fungi and bacteria. MAP kinase 4 in Arabidopsis is required to activate the salicylic acid and jasmonate pathways for defense response (Andreasson et al., 2005). In their study, Asai et al. (2002) discovered that when flagellin of bacteria is detected by Arabidopsis, it started a complete MAP kinase signaling cascade that activates the immunity.

The significant marker found in Pv05 (Pv05_30232782) was 12kb near the candidate gene Phvul.005G097400 that encodes a GTP-binding protein, HflX. GTP-binding proteins are GTPases involved in many cellular processes in plants (Urano et al., 2013). HflX family belongs to the translation factors (TRAFAC) group, which was one of the last groups of G-proteins that have been discovered (Verstraeten et al., 2011). The function of the TRAFAC family is mostly unknown. However, a recent study discovered that two GTP-binding proteins belonging to this family are involved in activating immunity against pathogens. When these genes were silenced, resistance of *N. benthamiana* to *P. syringae* was compromised (Lee et al., 2017). Similar to rust resistance, no R genes are reported, but genes related to PTI were found in the genomic regions.

Downy mildew resistance

Downy mildew in common bean is caused by *Phytophthora nicotianae* Breda de Haan var. *parasitica* (Dastur) G. M. Waterhouse or *Phytophthora phaseoli* Thaxt. *P. phaseoli*, which is mostly related to downy mildew in lima beans. *P. nicotianae* is mostly found in Latin America

and West Indies. No resistant genes for downy mildew have been reported (Schwartz et al., 2005).

In Pv04, significant markers Pv04_2305784 and Pv04_2305787 were located 12kb near to a gene model that encodes a cellulose-synthase-like C4 protein. Cell wall in plants is formed mostly by cellulose, it maintains the structure of the cell and constitute a barrier against pathogens, and it also communicates with the exterior using signaling proteins (Carpita and McCann, 2000). Cellulose synthases are in charge of the synthesis of cellulose for the primary and secondary cell walls (Hernandez-Blanco et al., 2007). Evidence suggests that mutations of the cellulose synthases can trigger the signaling pathways that enhance the resistance to pathogens in Arabidopsis, such as powdery mildew (*Erysiphe cichoracearum*) (Ellis et al., 2002).

In Pv06, two significant markers fell within the candidate gene Phvul.006G165300 that encodes a cullin 4 protein. Cullin 4, known as CUL4 is a subunit of the ubiquitin protein ligase complex (one of the enzymes needed by ubiquitin) involved in ubiquitination and degradation of proteins in the ubiquitin/26S pathway (Sullivan et al., 2003). As discussed above for seed shape and rust resistance, it has been demonstrated that ubiquitination is involved in many processes of the plant cell. Ligases needed for ubiquitination seems to have also a role in disease resistance mediated by R genes. It is suggested that the COP9 signalosome (CSN), another protein related to the ubiquitin/26S pathway, interacts with the cullin and therefore regulates the ubiquitin protein ligase. However, it is not clear yet which proteins are targeted by the pathway during the defense response activation (Liu et al., 2002).

The gene model found 5kb close to the significant markers Pv06_26759529 and Pv06_26760100 was Phvul.006G164700, which encodes a protein of unknown function

(DUF538). Proteins belonging to this family have not been studied in detail. However, they are suggested to be the homolog of the bactericidal/permeability increasing proteins in mammals (Gholizadeh and Kohnehrouz, 2013). DUF538 was proposed as a candidate protein that activates the plant response to abiotic and biotic stresses (Gholizadeh 2011). However, significant markers found in this region were located less than 100kb from the gene encoding CUL4 that was already described.

A last significant SNP in Pv06 was located almost 6 Mb apart from the other SNPs found in this chromosome. The region surrounding the marker has no annotated gene models. However, the candidate gene Phvul.006G102300 located 28 Kb from the marker encodes a pathogenesis-related (PR) 4 protein. PR proteins were first described in the 1970's in *Nicotiana tabacum* when evaluated under pressure of the tobacco mosaic virus. Nowadays, PR proteins are known to be within the most abundant proteins found in plants during defense responses (Breen et al., 2017). PR4 protein has been identified in the plasma membrane in *Capsicum annuum* when HR is activated by the pathogen *Xanthomonas campestris* effector AvrBsT (Kim and Hwang, 2015). It is suggested that PR4 has a function similar to the R genes. However, how the protein activates the defense response downstream needs to be elucidated (Kim and Hwang, 2015). The Guatemalan climbing beans can be a good source to evaluate candidate genes for this disease.

CONCLUSIONS

The genetic diversity and population structure analysis of the Guatemalan climbing bean collections allowed the confirmation of race Guatemala in the Middle American gene pool and its differentiation from race Mesoamerica and race Durango-Jalisco. Therefore, it represents a new source of alleles for breeding programs. It is recommended that seeds of the Guatemala populations GUA_1966-82 and GUA_2015 should be retained in the germplasm bank both among and within accessions. Single seed descendant is recommended before starting the pre-breeding process.

The lower population structure found within the Guatemalan climbing beans also makes this population ideal for the discovery of genetic factors and candidate genes for important traits such as altitudinal adaptation, seed shape, and factors affecting the color expression in beans. In this study, we demonstrate that the population was useful to provide candidate genes for previous reported genetic factors like the *V* gene for flower color, the *Prp* locus for pod color, and the *Asp* gene for seed coat luster. It was also useful to emphasize the important relationship between flowering time and altitudinal adaptation of common bean. This are the first steps for the identification of candidate genes. However, bi-parental crosses and fine mapping are necessary for the identification of the genes.

REFERENCES

- Acosta-Gallegos, J.A., J.D. Kelly, and P. Gepts. 2007. Prebreeding in common bean and use of genetic diversity from wild germplasm. *Crop Sci.* 47: S-44–S-59.
- Agarwal, M., N. Shrivastava, and H. Padh. 2008. Advances in molecular marker techniques and their applications in plant sciences. *Plant Cell Rep.* 27: 617-631.
- Agreda, K.A., J.C. Villatoro, A.N. Miranda, J.R. Moscoso, J.M. Osorno, and P. McClean. 2017. Phenotypic evaluation of native accessions of climbing beans collected in the Guatemalan highlands. Poster session presented at: Legume Innovation Lab Grain Legume Research Conference. Ouagadougou, Burkina Faso.
- Akibode, S., and M.K. Maredia. 2012. Global and regional trends in production, trade and consumption of food legume crops. Available at <https://impact.cgiar.org/sites/default/files/images/Legumetrendsv2.pdf> (accessed 10 October 2017)
- Alabadí, D., T. Oyama, M.J. Yanovsky, F.G. Harmon, P. Más, and S.A. Kay. 2001. Reciprocal regulation between TOC1 and LHY/CCA1 within the *Arabidopsis* circadian clock. *Science* 293: 880-883.
- Aldana, L.F. 2010. Manual of commercial production of common bean seed *Phaseolus vulgaris* (In Spanish). Instituto de Ciencia y Tecnología Agrícolas (ICTA). Guatemala. p. 37.
- Alzate-Marin, A.L., M.R. Costa, K.M. Arruda, E.G. de Barros, and M.A. Moreira. 2003. Characterization of the anthracnose resistance gene present in Ouro Negro (Honduras 35) common bean cultivar. *Euphytica* 133(2): 165–169.
- Andreasson, E., T. Jenkins, P. Brodersen, S. Thorgrimsen, N.H.T. Petersen, S. Zhu, J.L. Qiu, P. Micheelsen, A. Rocher, M. Petersen, M.A. Newman, H. Bjørn Nielsen, H. Hirt, I. Somssich, O. Mattsson, and J. Mundy. 2005. The MAP kinase substrate MKS1 is a regulator of plant defense responses. *EMBO J.* 24: 2579-2589.
- Asai, T., G. Tena, J. Plotnikova, M.R. Willmann, W.L. Chiu, L. Gomez-Gomez, T. Boller, F.M. Ausubel, and J. Sheen. 2002. MAP kinase signalling cascade in *Arabidopsis* innate immunity. *Nature* 415: 977-983.
- Augustin, E., D.P. Coyne, and M.L. Schuster. 1972. Inheritance of resistance in *Phaseolus vulgaris* to *Uromyces phaseoli typica* Brazilian rust race B11 and of plant habit. *J. Am. Soc. Hortic. Sci.* 97: 526–529.
- Aulchenko, Y.S., S. Ripke, A. Isaacs, and C.M. van Duijn. 2007. GenABEL: an R package for genome-wide association analysis. *Bioinformatics* 23: 1294-1296.
- Ballantyne, B. 1978. The genetic bases of resistance to rust, caused by *Uromyces Appendiculatus* in bean (*Phaseolus Vulgaris*). University of Sydney, Australia.
- Bannerot, H., M. Derieux, and G. Fouilloux. 1971. Mise en evidence d'un second gene de resistance totale a l'anthracnose chez le haricot. *Ann Amélior Plantes* 21: 83-85.
- Baskin, T.I. 2005. Anisotropic expansion of the plant cell wall. *Annu. Rev. Cell Dev. Biol.* 21: 203-222.
- Bassett, M.J. 1996. The margo (*mar*) seedcoat color gene is a synonym for the Joker (*j*) locus in common bean. *J. Am. Soc. Hortic. Sci.* 121: 1028-1031.
- Bassett, M.J. 2005. A new gene (*Prpi-2*) for intensified anthocyanin expression (IAE) syndrome in common bean and a reconciliation of gene symbols used by early investigators for purple pod and IAE syndrome. *J. Am. Soc. Hortic. Sci.* 130: 550-554.

- Bassett, M.J., X. Lin-Bao, and L.C. Hannah. 1990. Flower colors in common bean produced by interactions of the *Sal* and *V* loci and a gametophyte factor *Ga* linked to *Sal*. *J. Am. Soc. Hortic. Sci.* 115: 1029-1033.
- Beebe, S., P.W. Skroch, J. Tohme, M.C. Duque, F. Pedraza, and J. Nienhuis. 2000. Structure of genetic diversity among common bean landraces of Middle American origin based on correspondence analysis of RAPD. *Crop Sci.* 40: 264–273.
- Beninger, C.W., G.L. Hosfield, and M.J. Bassett. 1999. Flavonoid composition of three genotypes of dry bean (*Phaseolus vulgaris*) differing in seed coat color. *J. Am. Soc. Hortic. Sci.* 124: 514-518.
- Beninger, C.W., G.L. Hosfield, M.J. Bassett, and S. Owens. 2000. Chemical and morphological expression of the *B* and *Asp* seedcoat genes in *Phaseolus vulgaris*. *J. Am. Soc. Hortic. Sci.* 125: 52-58.
- Bhatramakki, D., M. Dolan, M. Hanafey, R. Wineland, D. Vaske, J.C. Register, S.V. Tingey, and A. Rafalski. 2002. Insertion-deletion polymorphisms in 3' regions of maize genes occur frequently and can be used as highly informative genetic markers. *Plant Mol. Biol.* 48: 539-547.
- Bitocchi, E., L. Nanni, E. Belluci, M. Rossi, A. Giardini, P.S. Zeuli, G. Logozzo, J. Stougaard, P. McClean, G. Attene, and R. Papa. 2012. Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proc. Natl. Acad. Sci. U. S. A.* 109: E788-E796.
- Blair, M.W., A.J. Cortés, R.V. Penmetsa, A. Farmer, N. Carrasquilla-García, and D.R. Cook. 2013. A high-throughput SNP marker system for parental polymorphism screening, and diversity analysis in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genetics* 126: 535–548.
- Blair, M.W., L.M. Díaz, H.F. Buendía, and M.C. Duque. 2009. Genetic diversity, seed size associations and population structure of a core collection of common beans (*Phaseolus vulgaris* L.). *Theor. Appl. Genetics* 119: 955-972.
- Blair, M.W., M.C. Giraldo, H.F. Buendia, E. Tovar, M.C. Duque, and S.E. Beebe. 2006. Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genetics* 113: 100-109.
- Blümel, M., N. Dally, and C. Jung. 2015. Flowering time regulation in crops-what did we learn from *Arabidopsis*? *Curr. Opin. Biotechnol.* 32: 121-129.
- Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss, and E.S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23: 2633-2635.
- Breen, S., S.J. Williams, M. Outram, B. Kobe, and P.S. Solomon. 2017. Emerging insights into the functions of Pathogenesis-related protein 1. *Trends Plant Sci.* 22: 871-879.
- Broughton, W.J., G. Hernández, M. Blair, S. Beebe, P. Gepts, and J. Vanderleyden. 2003. Beans (*Phaseolus spp.*) – model food legumes. *Plant Soil* 252: 55–128.
- Brown, J.K.M. 2003. A cost of disease resistance: paradigm or peculiarity? *Trends Genet.* 19: 667-671.
- Cao, L., J.L. Henty-Ridilla, L. Blanchoin, and C.J. Staiger. 2016. Profilin-dependent nucleation and assembly of actin filaments controls cell elongation in *Arabidopsis*. *Plant Physiol.* 170: 220-233.

- Cardenas, F., M.W. Adams, and A. Andersen. 1964. The genetic system for reaction of field beans (*Phaseolus vulgaris* L.) to infection by three physiologic races of *Colletotrichum lindemuthianum*. *Euphytica* 13(2): 178–186.
- Carpita, N.C., and M. McCann. 2000. The cell wall. p. 52–108. *In* Buchanan, B., W. Gruissem, and R.L. Jones, (eds), *Biochemistry and Molecular Biology of Plants*. American Society of Plant Physiologists. Rockville, MD.
- Chacón, 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. Genetics* 110: 432-444.
- Chen, K.Q., X.Y. Zhao, X.H. An, Y. Tian, D.D. Liu, C.X. You, and Y.J. Hao. 2017. MdHIR proteins repress anthocyanin accumulation by interacting with the MdJAZ2 protein to inhibit its degradation in apples. *Sci. Rep.* 7: 44484. doi:10.1038/srep44484.
- Chen, M., J. Wu, L. Wang, X. Zhang, M. Blair, J. Jia, and S. Wang. 2013. Development of mapped simple sequence repeat markers from common bean (*Phaseolus vulgaris* L.) based on genome sequences of a Chinese landrace and diversity evaluation. *Mol. Breeding* 33: 489-496.
- Cheng, Y., N. Kato, W. Wang, J. Li, and X. Chen. 2003. Two RNA binding proteins, HEN4 and HUA1, act in the processing of AGAMOUS pre-mRNA in *Arabidopsis thaliana*. *Dev. Cell.* 4: 53-66.
- Chisholm, S.T., G. Coaker, B. Day, and B.J. Staskawicz. 2006. Host-microbe interactions: shaping the evolution of the plant immune response. *Cell* 124: 803-814.
- Cichy, K.A., A. Fernandez, A. Kilian, J.D. Kelly, C.H. Galeano, S. Shaw, M. Brick, D. Hodkinson, and E. Troxtell. 2014. QTL analysis of canning quality and color retention in black beans (*Phaseolus vulgaris* L.). *Mol. Breed.* 33: 139-154.
- Cichy, K.A., T.G. Porch, J.S. Beaver, P. Cregan, D. Fourie, R.P. Glahn, M.A. Grusak, K. Kamfwa, D.N. Katuramu, P. McClean, and E. Mndolwa. 2015. A diversity panel for Andean bean improvement. *Crop Sci.* 55: 2149-2160.
- Clavijo, P. 1980. General summary of the main agronomic features of different grains in Colombia (In Spanish). Instituto Interamericano de Ciencias Agrícolas (IICA). Colombia. p. 49.
- Deleris, A., J. Gallego-Bartolome, J. Bao, K.D. Kasschau, J.C. Carrington, and O. Voinnet. 2006. Hierarchical action and inhibition of plant Dicer-like proteins in antiviral defense. *Science* 313: 68-71.
- Delgado-Salinas, A., T. Turley, A. Richman, and M. Lavin. 1999. Phylogenetic analysis of the cultivated and wild species of *Phaseolus* (Fabaceae). *Syst. Bot.* 1: 438-460.
- DePristo, M., E. Banks, R. Poplin, K. Garimella, J. Maguire, C. Hartl, A. Philippakis, G. del Angel, M.A. Rivas, M. Hanna, A. McKenna, T. Fennell, A. Kernysky, A. Sivachenko, K. Cibulskis, S. Gabriel, D. Altshuler, and M. Daly. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genet.* 43: 491-498.
- De Souza, T.L.P.O., S.N. Dessaune, D.A. Sanglard, M.A. Moreira, and E.G. de Barros. 2011. Characterization of the rust resistance gene present in the common bean cultivar Ouro Negro, the main rust resistance source used in Brazil. *Plant Pathol.* 60: 839–845.
- DeYoung, D.J., B. Reyes, J.M. Osorno, J.C. Villatoro, and M.K. Maredia. 2017. An overview of bean production practices, varietal preferences, and consumption patterns in the milpa system of the Guatemalan highlands: results of a farm household survey. Department of

- Agricultural, Food and Resource Economics, Michigan State University. Staff Paper No. 2017-08.
- Diamant, R., B.M. Watts, L.G. Elias, and B. Rios. 1989. Consumer utilization and acceptability of raw and cooked black beans (*Phaseolus Vulgaris*) in Guatemala. *Ecol. Food Nutr.* 22: 183-195.
- Diaz, L. M., and M.W. Blair. 2006. Race structure within the Mesoamerican gene pool of common bean (*Phaseolus vulgaris* L.) as determined by microsatellite markers. *Theor. Appl. Genetics* 114: 143-154.
- Du, L., N. Li, L. Chen, Y. Xu, Y. Li, Y. Zhang, C. Li, and Y. Li. 2014. The ubiquitin receptor DA1 regulates seed and organ size by modulating the stability of the ubiquitin-specific protease UBP15/SOD2 in *Arabidopsis*. *Plant Cell* 26: 665-677.
- Ebel, R., J.G. Pozas Cárdenas, F. Soria Miranda, and J. Cruz González. 2017. Organic milpa: yields of maize, beans, and squash in mono- and polycropping systems. (In Spanish). *Terra Latinoamericana* 35: 149-160.
- Ellis, C., I. Karafyllidis, C. Wasternack, and J.G. Turner. 2002. The *Arabidopsis* mutant *cevl* links cell wall signaling to jasmonate and ethylene responses. *Plant Cell* 14: 1557-1566.
- Elshire, R.J., J.C. Glaubitz, Q. Sun, J.A. Poland, K. Kawamoto, E.S. Buckler, and S.E. Mitchell. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6: e19379. doi:10.1371/journal.pone.0019379.
- Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14: 2611-2620.
- Ewan, R., R. Pangestuti, S. Thornber, A. Craig, C. Carr, L. O'Donnell, C. Zhang, and A. Sadanandom. 2011. Deubiquitinating enzymes AtUBP12 and AtUBP13 and their tobacco homologue NtUBP12 are negative regulators of plant immunity. *New Phytol.* 191: 92-106.
- FastQC. Babraham bioinformatics. Available at <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (accessed 10 May 2016).
- FASTX. FASTX-Toolkit-Hannon Lab. Available at http://hannonlab.cshl.edu/fastx_toolkit/ (accessed 10 May 2016).
- Ferreira, J.J., A. Campa, and J.D. Kelly. 2013. Organization of genes conferring resistance to Anthracnose in Common Bean. p. 151-181. *In* Translational Genomics for Crop Breeding. John Wiley & Sons Ltd.
- Figtree. 2007. Available at <http://tree.bio.ed.ac.uk/software/figtree/> (accessed 15 June 2017).
- Finke, M.L., D.P. Coyne, and J.R. Steadman. 1986. The inheritance and association of resistance to rust, common bacterial blight, plant habit and foliar abnormalities in *Phaseolus vulgaris* L. *Euphytica* 35: 969-982.
- Flor, H.H. 1955. Host-parasite interactions in flax rust-its genetics and other implications. *Phytopathology* 45: 680-685
- Fouilloux, G. 1976. L'Anthracnose du haricot (*Colletotrichum lindemuthianum* Sacc. et Magn.): Nouvelles sources de résistance et nouvelles races physiologiques. *Ann. Amélior Plantes* 26: 443-453.
- Franks, R.G., C. Wang, J.Z. Levin, and Z. Liu. 2002. SEUSS, a member of a novel family of plant regulatory proteins, represses floral homeotic gene expression with LEUNIG. *Development* 129: 253-263.
- Galeano, C.H., A.J. Cortes, A.C. Fernández, A. 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:<https://doi.org/10.1186/1471-2156-13-48>.
- Galvan, M.Z., B. Bornet, P.A. Balatti, and M. Branchard. 2003. Inter simple sequence repeat (ISSR) markers as a tool for the assessment of both genetic diversity and gene pool origin in common bean (*Phaseolus vulgaris* L.). *Euphytica* 132: 297-301.
- Geffroy, V., D. Sicard, J.C. de Oliveira, M. Sévignac, S. Cohen, P. Gepts, C. Neema, T. Langin, and M. Dron. 1999. Identification of an ancestral resistance gene cluster involved in the coevolution process between *Phaseolus vulgaris* and its fungal pathogen *Colletotrichum lindemuthianum*. *Mol. Plant. Microbe. Interact.* 12(9): 774–784.
- Geffroy, V., M. Sévignac, P. Billant, M. Dron, and T. Langin. 2008. Resistance to *Colletotrichum lindemuthianum* in *Phaseolus vulgaris*: a case study for mapping two independent genes. *Theor. Appl. Genet.* 116(3): 407–415.
- Gepts, P., and D. Debouck. 1991. Origin, domestication and evolution of the common bean (*Phaseolus vulgaris*L.). p. 7–53. In: Van Shoonhoven, A., and O. Voysest. *Common beans: Research for crop improvement*. (eds): Wallingford, Commonwealth Agricultural Bureau.
- 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.
- Gholizadeh, A. 2011. Heterologous expression of stress-responsive DUF538 domain containing protein and its morpho-biochemical consequences. *Protein J.* 30: 351-358.
- Gholizadeh, A., and S.B. Kohnhrouz. 2013. DUF538 protein super family is predicted to be the potential homologue of bactericidal/permeability-increasing protein in plant system. *Protein J.* 32: 163-171.
- Gonçalves-Vidigal, M.C., A.A. Cardoso, C. Vieira, and L.S. Saraiva. 1997. Inheritance of anthracnose resistance in common bean genotypes PI 207262 and AB 136. *Braz. J. of Genetics* 20: 59-62.
- Gonçalves-Vidigal, M.C., G.F. Lacanallo, and P.S. Vidigal Filho. 2008. A new gene conferring resistance to anthracnose in Andean common bean (*Phaseolus vulgaris* L.) cultivar “Jalo Vermelho.” *Plant Breed.* 127(6): 592–596.
- Grafton, K.F., G.C. Weiser, L.J. Littlefield, and J.R. Stavely. 1985. Inheritance of resistance to two races of leaf rust in dry edible bean. *Crop Sci.* 25: 537–539.
- Guo, J., W. Zhou, Z. Lu, H. Li, H. Li, and F. Gao. 2015. Isolation and functional analysis of chalcone isomerase gene from purple-fleshed sweet potato. *Plant Mol. Biol. Rep.* 33: 1451-1463.
- Gupta, P.K., S. Rustgi, and P.L. Kulwal. 2005. Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Mol. Biol.* 57: 461-485.
- Hernández-Blanco, C., D.X. Feng, J. Hu, A. Sánchez-Vallet, L. Deslandes, F. Llorente, M. Berrocal-Lobo, H. Keller, X. Barlet, C. Sánchez-Rodríguez, L.K. Anderson, S. Somerville, Y. Marco, and A. Molina. 2007. Impairment of cellulose synthases required for *Arabidopsis* secondary cell wall formation enhances disease resistance. *Plant Cell* 19: 890-903.
- Hjort, K., I. Presti, A. Elväng, F. Marinelli, and S. Sjöling. 2014. Bacterial chitinase with phytopathogen control capacity from suppressive soil revealed by functional metagenomics. *Appl. Microbiol. Biotechnol.* 98: 2819-2828.

- Hückelhoven, R., R. Eichmann, C. Weis, C. Hoefle, and R.K. Proels. 2013. Genetic loss of susceptibility: a costly route to disease resistance? *Plant Pathol.* 62: 56-62.
- Hughes, A.R., B.D. Inouye, M.T. Johnson, N. Underwood, and M. Vellend. 2008. Ecological consequences of genetic diversity. *Ecol. Lett.* 11: 609-623.
- Hurtado-Gonzales, O.P., G. Valentini, T.A.S. Gilio, A.M. Martins, Q. Song, and M.A. Pastor-Corrales. 2017. Fine mapping of Ur-3, a historically important rust resistance locus in common bean. *G3.* 7: 557-569.
- IICA. 2014. The value chains of white corn and common bean in Central America (In Spanish). San José, C.R. p. 127.
- Illumina.com-Introduction to DNA sequencing. Available at <http://www.illumina.com/techniques/sequencing/dna-sequencing.html> (accessed 09 May 2016).
- Jeong, S.T., N. Goto-Yamamoto, S. Kobayashi, and M. Esaka. 2004. Effects of plant hormones and shading on the accumulation of anthocyanins and the expression of anthocyanin biosynthetic genes in grape berry skins. *Plant Sci.* 167: 247-252.
- Joshi, J., B. Schmid, M.C. Caldeira, P.G. Dimitrakopoulos, J. Good, R. Harris, A. Hector, K. Huss-Danell, A. Jumpponen, A. Minns, C.P.H. Mulder, J.S. Pereira, A. Prinz, M. Scherer-Lorenzen, A.S.D. Siamantziouras, A.C. Terry, A.Y. Troumbis, and J.H. Lawton. 2001. Local adaptation enhances performance of common plant species. *Ecol. Lett.* 4: 536-544.
- Joshi, N.A., and J.N. Fass. 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at <https://github.com/najoshi/sickle> (accessed 10 May 2016).
- Jung, G., D.P. Coyne, J. Bokosi, J.R. Steadman, and J. Nienhuis. 1998. Mapping genes for specific and adult plant resistance to rust and abaxial leaf pubescence and their genetic relationships using randomly amplified polymorphic DNA (RAPD) markers in common bean. *J. Am. Soc. Hortic. Sci.* 123: 859-863.
- Jurado, S., S.D. Triviño, Z. Abraham, C. Manzano, C. Gutierrez, and C. Del Pozo. 2008. SKP2A protein, an F-box that regulates cell division, is degraded via the ubiquitin pathway. *Plant Signal. Behav.* 3: 810-812.
- Kami, J., V. Becerra-Velasquez, D. Debouck, and P. Gepts. 1995. Identification of presumed ancestral DNA sequences of phaseolin in *Phaseolus vulgaris*. *Proc. Natl. Acad. Sci. U. S. A.* 92: 1101-1104.
- Kang, H.M., N.A Zaitlen, C.M. Wade, A. Kirby, D. Heckerman, M.J. Daly, and E. Eskin. 2008. Efficient control of population structure in model organism association mapping. *Genetics* 178: 1709-1723.
- Kaplan, L. 1965. Archeology and domestication in American *Phaseolus* (Beans). *Econ. Bot.* 19: 358-368.
- Kelly, J.D. 2010. The story of bean breeding: Michigan State University publication prepared for BeanCAP & works, PBG.
- Kelly, J.D., and V.A. Vallejo. 2004. A comprehensive review of the major genes conditioning resistance to Anthracnose in common bean. *Hort. Science* 39: 1196-1207.
- Kessler, S.A., H. Shimosato-Asano, N.F. Keinath, S.E. Wuest, G. Ingram, R. Panstruga, and U. Grossniklaus. 2010. Conserved molecular components for pollen tube reception and fungal invasion. *Science* 330: 968-971.

- Kim, N.H., and B.K. Hwang. 2015. Pepper pathogenesis-related protein 4c is a plasma membrane-localized cysteine protease inhibitor that is required for plant cell death and defense signaling. *Plant J.* 81: 81-94.
- Kirsch, C., E. Logemann, B. Lippok, E. Schmelzer, and K. Hahlbrock. 2001. A highly specific pathogen-responsive promoter element from the immediate-early activated CMPG1 gene in *Petroselinum crispum*. *Plant J.* 26: 217-227.
- Klaedtke, S.M., L. Caproni, J. Klauck, P. de la Grandville, M. Dutartre, P.M. Stassart, V. Chable, V. Negri, and L. Raggi. 2017. Short-term local adaptation of historical common bean (*Phaseolus vulgaris* L.) varieties and implications for in situ management of bean diversity. *Int. J. Mol. Sci.* 18:493-511.
- Konzen, E.R., and S.M. Tsai. 2014. Seed coat shininess in *Phaseolus vulgaris*: rescuing a neglected trait by its screening on commercial lines and landraces. *J. Agric. Sci.* 6: 113-130.
- Kumar, S., G. Stecher, and K. Tamura. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33(7): 1870–1874
- 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. Genetics* 118: 979–992.
- Lachniet, M.S., and W.P. Patterson. 2009. Oxygen isotope values of precipitation and surface waters in northern Central America (Belize and Guatemala) are dominated by temperature and amount effects. *Earth Planet. Sci. Lett.* 284: 435–446.
- Lam, B.C.H., T.L. Sage, F. Bianchi, and E. Blumwald. 2001. Role of SH3 domain-containing proteins in clathrin-mediated vesicle trafficking in *Arabidopsis*. *Plant Cell* 13: 2499-2512.
- Lamprecht, H. 1940. Zur genetik von *Phaseolus vulgaris* XVII-XVIII. *Hereditas* 26: 292-304.
- Lee, S., M. Senthil-Kumar, M. Kang, C.M. Rojas, Y. Tang, S. Oh, S.R. Choudhury, H.K. Lee, Y. Ishiga, R.D. Allen, S. Pandey, and K.S. Mysore. 2017. The small GTPase, nucleolar GTP-binding protein 1 (NOG1), has a novel role in plant innate immunity. *Sci. Rep.* 7: 9260. doi:10.1038/s41598-017-08932-9.
- Lee, T.H., H. Guo, X. Wang, C. Kim, and A.H. Paterson. 2014. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics* 15:162. doi:https://doi.org/10.1186/1471-2164-15-162
- Legume Innovation Lab (LIL). 2017. Two new climbing bean varieties offer hope for smallholder farmers in Guatemala. Available at http://legumelab.msu.edu/uploads/files/LIL_Project_Releases_Two_New_Bean_Varieties_SS2017.pdf (accessed 04 Dec. 2017).
- Lenhard, M., A. Bohnert, G. Jürgens, and T. Laux. 2001. Termination of stem cell maintenance in *Arabidopsis* floral meristems by interactions between WUSCHEL and AGAMOUS. *Cell* 105: 805-814.
- Lepiniec, L., I. Debeaujon, J.M. Routaboul, A. Baudry, L. Pourcel, N. Nesi, and M. Caboche. 2006. Genetics and biochemistry of seed flavonoids. *Annu. Rev. Plant Biol.* 57: 405-430.
- Lépiz-Ildefonso, R., J.J. López-Alcocer, J.J. Sánchez-González, F. Santacruz-Ruvalcaba, R. Nuño-Romero, and E. Rodríguez-Guzmán. 2010. Morphological features of cultivated forms, wild types and intermediate forms of climbing common bean (In Spanish). *Revista Fitotecnia Mexicana* 33: 21-28.
- Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v1 [q-bio.GN].

- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078-2079.
- Li, H., Z. Peng, X. Yang, W. Wang, J. Fu, J. Wang, Y. Han, Y. Chai, T. Guo, N. Yang, and J. Liu. 2013. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nature Genet.* 45: 43-50.
- Li, T., J. Jiang, S. Zhang, H. Shu, Y. Wang, J. Lai, J. Du, and C. Yang. 2015. OsAGSW1, an ABC1-like kinase gene, is involved in the regulation of grain size and weight in rice. *J. Exp. Bot.* 66: 5691-5701.
- Liebenberg, M.M., and Z.A. Pretorius. 2004. Inheritance of resistance to *Uromyces appendiculatus* in the South African dry bean cultivar Kranskop. *S. Afr. J. Plant Soil./S.-Afr. Tydskr. Plant Grond* 21: 245–250.
- Liu, K., and S.V. Muse. 2005. PowerMarker. Integrated analysis environment for genetic marker data. *Bioinformatics* 21: 2128-2129.
- Liu, Y., M. Schiff, G. Serino, X.W. Deng, and S.P. Dinesh-Kumar. 2002. Role of SCF ubiquitin-ligase and the COP9 signalosome in the N gene-mediated resistance response to Tobacco mosaic virus. *Plant Cell* 14: 1483-1496.
- Loveless, M.D., and J.L. Hamrick. 1984. Ecological determinants of genetic structure in plant populations. *Annu. Rev. Ecology Systemat.* 1: 65-95.
- Luo, Q., Q. Wei, R. Wang, Y. Zhang, F. Zhang, Y. He, S. Zhou, J. Feng, G. Yang, and G. He. 2017. BdCIPK31, a calcineurin B-like protein-interacting protein kinase, regulates plant response to drought and salt stress. *Front. Plant Sci.* 8: 1184.
doi:10.3389/fpls.2017.01184
- Maldonado-Mota, C.R., M.A. Pastor-Corrales, O.P. Hurtado-González, S.M. Moghaddam, S. Schröder, P. McClean, J. Pasche, R. Lamppa, M.G. Tobar-Piñón, and J.M. Osorno. 2017. Identification of new sources of resistance to anthracnose in climbing bean germplasm from Guatemala. Oral session presented at: Bean Improvement Cooperative Conference. East Lansing MI, USA.
- Mamidi, S., M. Rossi, D. Annam, S. Moghaddam, R. Lee, R. Papa, and P. McClean. 2011. Investigation of the domestication of common bean (*Phaseolus vulgaris*) using multilocus sequence data. *Funct. Plant Biol.* 38: 953-967.
- Mamidi, S., M. Rossi, S.M. Moghaddam, D. Annam, R. Lee, R. Papa, and P.E. McClean. 2013. Demographic factors shaped diversity in the two gene pools of wild common bean *Phaseolus vulgaris* L. *Heredity* 110: 267–276.
- Mamidi, S., R.K. Lee, J.R. Goos, and P.E. McClean. 2014. Genome-wide association studies identifies seven major regions responsible for iron deficiency chlorosis in soybean (*Glycine max*). *PloS one* 9:e107469.
- 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. 2011b. Genome-wide association analysis identifies candidate genes associated with iron deficiency chlorosis in soybean. *Plant Genome* 4: 154–164.
- Mardis, E.R. 2008. The impact of next-generation sequencing technology on genetics. *Trends Genet.* 24: 133-141.
- Marini, A., and M. Gragnolati. 2003. Malnutrition and poverty in Guatemala. World Bank Policy Research Working Paper No. 2967. World Bank, Washington, DC.

- Mastenbroek, C. 1960. A breeding programme for resistance to anthracnose in dry shell haricot beans, based on a new gene. *Euphytica* 9: 177-84.
- Matuoka, K., F. Shibasaki, M. Shibata, and T. Takenawa. 1993. Ash/Grb-2, a SH2/SH3-containing protein, couples to signaling for mitogenesis and cytoskeletal reorganization by EGF and PDGF. *EMBO J.* 12: 3467-3473.
- McClellan, P.E., J. Terpstra, M. McConnell, C. White, R. Lee, and S. Mamidi. 2012. Population structure and genetic differentiation among the USDA common bean (*Phaseolus vulgaris* L.) core collection. *Genet. Resour. Crop Ev.* 59: 499-515.
- McClellan, P.E., R.K. Lee, C. Otto, P. Gepts, and M.J. Bassett. 2002. Molecular and phenotypic mapping of genes controlling seed coat pattern and color in common bean (*Phaseolus vulgaris* L.). *J. Hered.* 93: 148-152.
- McClellan, P.E., S.M. Moghaddam, A.F. Lopéz-Millán, M.A. Brick, J.D. Kelly, P.N. Miklas, J. Osorno, T.G. Porch, C.A. Urrea, A. Soltani, and M.A. Grusak. 2017. Phenotypic diversity for seed mineral concentration in North American dry bean germplasm of Middle American ancestry. *Crop Sci.* 57: 3129–3144.
- Mercati, F., M. Leone, A. Lupini, A. Sorgona, M. Bacchi, M.R. Abenavoli, and F. Sunseri. 2013. Genetic diversity and population structure of a common bean (*Phaseolus vulgaris* L.) collection from Calabria (Italy). *Genet. Res. Crop Evol.* 60: 839–852.
- Meziadi, C., M.M.S. Richard, A. Derquennes, V. Thareau, S. Blanchet, A. Gratias, S. Pflieger, and V. Geffroy. 2016. Development of molecular markers linked to disease resistance genes in common bean based on whole genome sequence. *Plant Sci.* 242: 351-357.
- Mielczarek, M., and J. Szyda. 2015. Review of alignment and SNP calling algorithms for next-generation sequencing data. *J. Appl. Genet.* 57: 71–79.
- Miklas, P.N., R. Delorme, V. Stone, M.J. Daly, J.R. Stavely, J.R. Steadman, M.J. Bassett, and J.S. Beaver. 2000. Bacterial, fungal, and viral disease resistance loci mapped in a recombinant inbred common bean population (Dorado'/XAN 176). *J. Am. Soc. Hortic. Sci.* 125: 476-481.
- Ministerio de Agricultura, Ganaderia y Alimentacion (MAGA). 2013. Prices of black bean: production of black bean. (In Spanish). Guatemala, Guatemala. p. 14.
- Moghaddam, S.M., Q. Song, S. Mamidi, J. Schmutz, R. Lee, P. Cregan, J.M. Osorno, and P.E. McClellan. 2014. Developing market class specific InDel markers from next generation sequence data in *Phaseolus vulgaris* L. *Front. Plant Sci.* 5:185.
- Moghaddam, S.M., S. Mamidi, J.M. Osorno, R. Lee, M. Brick, J. Kelly, P. Miklas, C. Urrea, Q. Song, P. Cregan, and J. Grimwood. 2016. Genome-wide association study identifies candidate loci underlying agronomic traits in a Middle American diversity panel of common bean. *Plant Genome* 9(3). doi:10.3835/plantgenome2016.02.0012.
- Montejo-Dominguez, L.M., J. Villatoro, J. Moscoso, J. Steadman, P. McClellan, and J.M. Osorno. 2017. Rust resistance in the Guatemalan climbing bean germplasm collection. Oral session presented at: Legume Innovation Lab Grain Legume Research Conference. Ouagadougou, Burkina Faso.
- Mouradov, A., F. Cremer, and G. Coupland. 2002. Control of flowering time: interacting pathways as a basis for diversity. *Plant Cell* 14 Suppl: S111–30. doi:<https://doi.org/10.1105/tpc.001362>
- Müller, B.S.F., G.J. Pappas, P.A.M.R. Valdisser, G.R.C. Coelho, I.P.P. de Menezes, A.G. Abreu, T.C.O. Borba, T. Sakamoto, C. Brondani, E.G. Barros, and R.P. Vianello. 2015. An operational SNP panel integrated to SSR marker for the assessment of genetic

- diversity and population structure of the common bean. *Plant Mol. Biol. Rep.* 33: 1697-1711.
- Muñoz, G., G. Giraldo, and J. Fernández de Soto. 1993. Varietal descriptors: rice, common bean, maize, sorghum (In Spanish). International Center for Tropical Agriculture (CIAT). Cali, Colombia. p. 174.
- Nambara, E., and A. Marion-Poll. 2005. Abscisic acid biosynthesis and catabolism. *Annu. Rev. Plant Biol.* 56: 165-185.
- Nei, M., and W.H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. U. S. A.* 76: 5269-5273.
- Nicaise, V., M. Roux, and C. Zipfel. 2009. Recent advances in PAMP-triggered immunity against bacteria: pattern recognition receptors watch over and raise the alarm. *Plant Physiol.* 150: 1638-1647.
- Niwa, Y., S. Ito, N. Nakamichi, T. Mizoguchi, K. Niinuma, T. Yamashino, and T. Mizuno. 2007. Genetic linkages of the circadian clock-associated genes, TOC1, CCA1 and LHY, in the photoperiodic control of flowering time in *Arabidopsis thaliana*. *Plant Cell Physiol.* 48: 925-937.
- Nordborg, M., and D. Weigel. 2008. Next-generation genetics in plants. *Nature* 456: 720-723.
- Okonkwo, C.A., and C.D. Clayberg. 1984. Genetics of flower and pod color in *Phaseolus vulgaris*. *J. Hered.* 75: 440-444.
- Olsen, K.M., and J.F. Wendel. 2013. A bountiful harvest: genomic insights into crop domestication phenotypes. *Annu. Rev. Plant Biol.* 64: 47-70.
- Olsen, K.M., U.S. Lea, R. Slimestad, M. Verheul, and C. Lillo. 2008. Differential expression of four *Arabidopsis* PAL genes; PAL1 and PAL2 have functional specialization in abiotic environmental-triggered flavonoid synthesis. *J. Plant Physiol.* 165: 1491-1499.
- Orellana, A., J. Villatoro, and M. Mérida. 2006. Morphological characterization and preliminary evaluation of the climbing bean germplasm collection (In Spanish). Instituto de Ciencia y Tecnología Agrícolas (ICTA). Guatemala, Guatemala. p.17.
- Osorno, J.M., and P. Mcclean. 2013. Genetic improvement of Middle-American climbing beans in Guatemala (SO1.A1). Feed the Future, Legume Innovation Lab. Michigan State University, Michigan. Available at http://legumelab.msu.edu/uploads/files/SO1.A1-FY2015_Annual_Technical_Progress_Report_Leg_Innovation_Lab.pdf (accessed 10 Oct. 2017).
- Pandey, G.K., P. Kanwar, A. Singh, L. Steinhorst, A. Pandey, A.K. Yadav, I. Tokas, S. Sanyal, B.G. Kim, S.C. Lee, Y.H. Cheong, J. Kudla, and S. Luan. 2015. CBL-interacting protein kinase, CIPK21, regulates osmotic and salt stress responses in *Arabidopsis*. *Plant Physiol.* 169: 780-92.
- Papa, R., and P. Gepts. 2003. Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from Mesoamerica. *Theor. Appl. Genetics* 106: 239-250.
- Peakall, R., and P.E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Ecol. Notes* 6: 288-295.
- Peakall, R., and P.E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* 28: 2537-2539.
- Picard tools. Broad Institute. Available at <http://broadinstitute.github.io/picard> (accessed 10 May 2016).

- Poczai, P., I. Varga, M. Laos, A. Cseh, N. Bell, J.P. Valkonen, and J. Hyvönen. 2013. Advances in plant gene-targeted and functional markers: a review. *Plant Methods* 9: 1-32.
- Poland, J.A., P.J. Brown, M.E. Sorrells, and J.L. Jannink. 2012. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PloS one* 7: p.e32253.
- Ponciano-Samayoa, K., J. Villatoro-Mérida, and L. Molina. 2009. Preliminary microsatellite characterization of Guatemalan climbing beans. (In Spanish). *Agronomía Mesoamericana* 20: 245-254.
- Porch, T., J. Beaver, D. Debouck, S. Jackson, J. Kelly, and H. Dempewolf. 2013. Use of wild relatives and closely related species to adapt common bean to climate change. *Agronomy* 3: 433–461.
- Prakken, R. 1972. Inheritance of colours in *Phaseolus vulgaris* L. III. On genes for red seed coat colour and a general synthesis. *Meded Landbouwhogeschool Wageningen* 29:1–82.
- Pritchard, J.P., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.
- Qi, P., Y.S. Lin, X.J. Song, J.B. Shen, W. Huang, J.X. Shan, M.Z. Zhu, L. Jiang, J.P. Gao, and H.X. Lin. 2012. The novel quantitative trait locus GL3.1 controls rice grain size and yield by regulating Cyclin-T1;3. *Cell Res.* 22: 1666-1680.
- Qi, T., S. Song, Q. Ren, D. Wu, H. Huang, Y. Chen, M. Fan, W. Peng, C. Ren, and D. Xie. 2011a. The Jasmonate-ZIM-domain proteins interact with the WD-Repeat/bHLH/MYB complexes to regulate Jasmonate-mediated anthocyanin accumulation and trichome initiation in *Arabidopsis thaliana*. *Plant Cell* 23: 1795-1814.
- Qi, Y., K. Tsuda, J. Glazebrook, and F. Katagiri. 2011b. Physical association of pattern-triggered immunity (PTI) and effector-triggered immunity (ETI) immune receptors in *Arabidopsis*. *Mol. Plant Pathol.* 12: 702-708.
- Rafiqi, M., M. Bernoux, J.G. Ellis, and P.N. Dodds. 2009. In the trenches of plant pathogen recognition: Role of NB-LRR proteins. *Semin. Cell Dev. Biol.* 20(9): 1017–1024.
- Ramachandran, S., H.E. Christensen, Y. Ishimaru, C.H. Dong, W. Chao-Ming, A.L. Cleary, and N.H. Chua. 2000. Profilin plays a role in cell elongation, cell shape maintenance, and flowering in *Arabidopsis*. *Plant Physiol.* 124: 1637-1647.
- R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL <http://www.R-project.org/>.
- Rendón-Anaya, M., J.M. Montero-Vargas, S. Saburido-Álvarez, A. Vlasova, S. Capella-Gutierrez, J.J. Ordaz-Ortiz, O.M. Aguilar, R.P. Vianello-Brondani, M. Santalla, L. Delaye, and T. Gabaldón. 2017. Genomic history of the origin and domestication of common bean unveils its closest sister species. *Genome Biol.* 18:60.
- Rodriguez, M., D. Rau, E. Bitocchi, E. Bellucci, E. Biagetti, A. Carboni, P. Gepts, L. Nanni, R. Papa, and G. Attene. 2016. Landscape genetics, adaptive diversity and population structure in *Phaseolus vulgaris*. *New Phytol.* 209: 1781-1794.
- Romero Navarro, J.A., M. Willcox, J. Burgueño, C. Romay, K. Swarts, S. Trachsel, E. Preciado, A. Terron, H.V. Delgado, V. Vidal, A. Ortega, A.E. Banda, N.O.G. Montiel, I. Ortiz-Monasterio, F.S. Vicente, A.G. Espinoza, G. Atlin, P. Wenzl, S. Hearne, and E.S. Buckler. 2017. A study of allelic diversity underlying flowering-time adaptation in maize landraces. *Nat. Genet.* 49: 476-480.
- Rosenberg, N.A. 2004. Distruct: A program for the graphical display of population structure. *Mol. Ecol. Notes* 4: 137-138.

- 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.
- 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. *Nat. Genet.* 46: 707–13.
- Schröder, S., S. Mamidi, R. Lee, M.R. McKain, P.E. McClean, and J.M. Osorno. 2016. Optimization of genotyping by sequencing (GBS) data in common bean (*Phaseolus vulgaris* L.). *Mol. Breed.* 36: 6.
- Schwartz, H. 2005. Anthracnose. *Compendium of bean diseases*, Second Edition. The American Phytopathological Society. USA. p.109.
- SESAN. Chronic malnutrition in Guatemala (In Spanish). Available at <http://www.sesan.gob.gt/index.php/incopas/item/1424-guatemala-reduce-desnutricion-segun-indice-global-del-hambre-elaborado-por-ifpri> (accessed 15 November 2015).
- Shan, X., Y. Zhang, W. Peng, Z. Wang, and D. Xie. 2009. Molecular mechanism for jasmonate-induction of anthocyanin accumulation in *Arabidopsis*. *J. Exp. Bot.* 60: 3849–3860.
- Shi, M.Z., and D.Y. Xie. 2014. Biosynthesis and metabolic engineering of anthocyanins in *Arabidopsis thaliana*. *Recent Pat. Biotechnol.* 8: 47-60.
- Sicard, D., P.S. Pennings, C. Grandclément, J. Acosta, O. Kaltz, and J.A. Shykoff. 2007. Specialization and local adaptation of a fungal parasite on two host plant species as revealed by two fitness traits. *Evolution* 61: 27-41.
- Singh, S. 1999. Common bean improvement in the twenty-first century. Kluwer, Dordrecht. The Netherlands. p.1–24.
- Singh, S.P., H. Terán, M. Lema, D.M. Webster, C.A. Strausbaugh, P.N. Miklas, H.F. Schwartz, and M.A. Brick. 2007. Seventy-five years of breeding dry bean of the Western USA. *Crop Sci.* 47: 981–989.
- Singh, S.P., P. Gept, and D. Debouck. 1991a. Races of common bean (*Phaseolus vulgaris*, Fabaceae). *Econ. Bot.* 45: 379-396.
- Singh, S.P., R. Nodari, and P. Gepts. 1991b. Genetic diversity in cultivated common bean: I. Allozymes. *Crop Sci.* 31: 19-23.
- Sirichandra, C., A. Wasilewska, F. Vlad, C. Valon, and J. Leung. 2009. The guard cell as a single-cell model towards understanding drought tolerance and abscisic acid action. *J. Exp. Bot.* 60: 1439-1463.
- Smalle, J., and R.D. Vierstra. 2004. The ubiquitin 26S proteasome proteolytic pathway. *Annu. Rev. Plant Biol.* 55: 555-590.
- Song, X.J., W. Huang, M. Shi, M.Z. Zhu, and H.X. Lin. 2007. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* 39: 623-630.

- Song, W.Y., G.L. Wang, L.L. Chen, H.S. Kim, L.Y. Pi, T. Holsten, J. Gardner, B. Wang, W.X. Zhai, L.H. Zhu, C. Fauquet, and P. Ronald. 1995. A receptor kinase-like protein encoded by the rice disease resistance gene, Xa21. *Science* 270: 1804-1806.
- Sonnante, G., T. Stockton, R.O. Nodari, V.B. Velásquez, and P. Gepts. 1994. Evolution of genetic diversity during the domestication of common-bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genetics* 89: 629-635.
- Stavely, J.R. 1984. Pathogenic specialization in *Uromyces phaseoli* in the United States and rust resistance in beans. *Plant Dis.* 68: 95-99.
- Stavely, J.R. 1990. Genetics of rust resistance in *Phaseolus vulgaris* plant introduction PI 181996. *Phytopathology* 80: 1056.
- Stavely, J.R., and M.A. Pastor-Corrales. 1989. Rust. p. 159-1942. *In: Schwartz, H.F., and M.A. Pastor-Corrales, (eds.). Bean production problems in the tropics. Centro Internacional de Agricultura Tropical (CIAT). Cali, CO.*
- Stegmann, M., J. Monaghan, E. Smakowska-Luzan, H. Rovenich, A. Lehner, N. Holton, Y. Belkhadir, and C. Zipfel. 2017. The receptor kinase FER is a RALF-regulated scaffold controlling plant immune signaling. *Science* 355: 287-289.
- Stracke, R., O. Jahns, M. Keck, T. Tohge, K. Niehaus, A.R. Fernie, and B. Weisshaar. 2010. Analysis of PRODUCTION OF FLAVONOL GLYCOSIDES-dependent flavonol glycoside accumulation in *Arabidopsis thaliana* plants reveals MYB11-, MYB12- and MYB111-independent flavonol glycoside accumulation. *New Phytol.* 188: 985-1000.
- Sullivan, J.A., K. Shirasu, and X.W. Deng. 2003. The diverse roles of ubiquitin and the 26S proteasome in the life of plants. *Nat. Rev. Genet.* 4: 948-958.
- Tanaka, Y., and F. Brugliera. 2013. Flower colour and cytochromes P450. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 368: 20120432. doi:10.1098/rstb.2012.0432.
- Umezawa, T., N. Sugiyama, M. Mizoguchi, S. Hayashi, F. Myouga, K. Yamaguchi-Shinozaki, Y. Ishihama, T. Hirayama, and K. Shinozaki. 2009. Type 2C protein phosphatases directly regulate abscisic acid-activated protein kinases in *Arabidopsis*. *Proc. Natl. Acad. Sci. U. S. A.* 106: 17588-17593.
- Urano, D., J.G. Chen, J.R. Botella, and A.M. Jones. 2013. Heterotrimeric G protein signalling in the plant kingdom. *Open Biol.* 3: 120186. doi:10.1098/rsob.120186.
- USDA-ERS. 2016. Dry edible beans. USDA-ERS. Available at <http://www.ers.usda.gov/topics/crops/vegetables-pulses/dry-beans.aspx> (accessed 25 January 2016).
- Väli, Ü., M. Brandström, M. Johansson, and H. Ellegren. 2008. Insertion-deletion polymorphisms (indels) as genetic markers in natural populations. *BMC Genetics* 9:8.
- Vandermarck, G.J., M.A. Brick, J.M. Osorno, J.D. Kelly, and C.A. Urrea. 2014. Edible grain legumes. p. 87-124. *In: Smith, S., B. Diers, J. Specht, and B. Carver, (eds). Yield gains in major U.S. field crops. ASA, CSSA, SSSA, Madison, WI.*
- Velásquez, V.L.B., and P. Gepts. 1994. RFLP diversity of common bean (*Phaseolus vulgaris*) in its centers of origin. *Genome* 37: 256-263.
- Verstraeten, N., M. Fauvart, W. Versées, and J. Michiels. 2011. The universally conserved prokaryotic GTPases. *Microbiol. Mol. Biol. Rev.* 75: 507-542.
- Vidali, L., R.C. Augustine, K.P. Kleinman, and M. Bezanilla. 2007. Profilin is essential for tip growth in the moss *Physcomitrella patens*. *Plant Cell* 19: 3705-3722.
- Vierstra, R.D. 2003. The ubiquitin/26S proteasome pathway, the complex last chapter in the life of many plant proteins. *Trends Plant Sci.* 8: 135-142.

- Vigouroux, Y., M. McMullen, C.T. Hittinger, K. Houchins, L. Schulz, S. Kresovich, Y. Matsuoka, and J. Doebley. 2002. Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. *Proc. Natl. Acad. Sci. U. S. A.* 99: 9650-9655.
- Vlasova, A., S. Capella-Gutiérrez, M. Rendón-Anaya, M. Hernández-Oñate, A.E. Minoche, I. Erb, F. Câmara, P. Prieto-Barja, A. Corvelo, W. Sanseverino, G. Westergaard, J.C. Dohm, G.J. Pappas Jr, S. Saburido-Alvarez, D. Kedra, I. Gonzalez, L. Cozzuto, J. Gómez-Garrido, M.A. Aguilar-Morón, N. Andreu, O.M. Aguilar, J. Garcia-Mas, M. Zehnsdorf, M.P. Vázquez, A. Delgado-Salinas, L. Delaye, E. Lowy, A. Mentaberry, R.P. Vianello-Brondani, J.L. García, T. Alioto, F. Sánchez, H. Himmelbauer, M. Santalla, C. Notredame, T. Gabaldón, A. Herrera-Estrella, and R. Guigó. 2016. Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes. *Genome Biol.* 17: 32.
- Wang, F., A. Muto, J. Van de Velde, P. Neyt, K. Himanen, K. Vandepoele, and M. Van Lijsebettens. 2015. Functional analysis of the *Arabidopsis* TETRASPANIN gene family in plant growth and development. *Plant Physiol.* 169: 2200-2214.
- Wang, N., Z. Zhang, S. Jiang, H. Xu, Y. Wang, S. Feng, and X. Chen. 2016. Synergistic effects of light and temperature on anthocyanin biosynthesis in callus cultures of red-fleshed apple (*Malus sieversii* f. *niedzwetzkyana*). *Plant Cell Tissue Organ Cult.* 127: 217-227.
- Xie, Z., E. Allen, A. Wilken, and J.C. Carrington. 2005. DICER-LIKE 4 functions in trans-acting small interfering RNA biogenesis and vegetative phase change in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* 102: 12984-12989.
- Yáñez-Mó, M., O. Barreiro, M. Gordon-Alonso, M. Sala-Valdés, and F. Sánchez-Madrid. 2009. Tetraspanin-enriched microdomains: a functional unit in cell plasma membranes. *Trends Cell Biol.* 19: 434-446.
- Yee, D., and D.R. Goring. 2009. The diversity of plant U-box E3 ubiquitin ligases: from upstream activators to downstream target substrates. *J. Exp. Bot.* 60: 1109-1121.
- Yuste-Lisbona, F.J., A.M. González, C. Capel, M. García-Alcázar, J. Capel, A.M. De Ron, M. Santalla, and R. Lozano. 2014. Genetic variation underlying pod size and color traits of common bean depends on quantitative trait loci with epistatic effects. *Mol. Breed.* 33: 939-952.
- Zhang, X., J. Wang, J. Huang, H. Lan, C. Wang, C. Yin, Y. Wu, H. Tang, Q. Qian, J. Li, and H. Zhang. 2012. Rare allele of OsPPKL1 associated with grain length causes extra-large grain and a significant yield increase in rice. *Proc. Natl. Acad. Sci. U. S. A.* 109: 21534-21539.
- Zhang, Z., E. Ersoz, C.Q. Lai, R.J. Todhunter, H.K. Tiwari, M.A. Gore, P.J. Bradbury, J. Yu, D.K. Arnett, J.M. Ordovas, and E.S. Buckler. 2010. Mixed linear model approach adapted for genome-wide association studies. *Nature Genet.* 42: 355-360.
- Zheng, J., Y. Zhang, and C. Wang. 2015. Molecular functions of genes related to grain shape in rice. *Breed. Sci.* 65: 120-126.
- Zhou, H., J. Zhao, J. Cai, and S.B. Patil. 2017. Ubiquitin-specific proteases function in plant development and stress responses. *Plant Mol. Biol.* 94: 565-576.

- Zhou, L., W. Lan, B. Chen, W. Fang, and S. Luan. 2015. A calcium sensor-regulated protein kinase, calcineurin b-like protein-interacting protein kinase19, is required for pollen tube growth and polarity. *Plant Physiol.* 167: 1351-1360.
- Zhu, C., M. Gore, E.S. Buckler, and J. Yu. 2008. Status and prospects of association mapping in plants. *Plant Genome* 1: 5-20.
- Zipfel, C. 2009. Early molecular events in PAMP-triggered immunity. *Curr. Opin. Plant Biol.* 12: 414-420.
- Zizumbo-Villarreal, D., and P. Colunga-García. 2012. The archaic diet in Mesoamerica: incentive for Milpa development and species domestication. *Econ. Bot.* 66: 328-43.

APPENDIX

Table A1. InDel Markers used in the intra-accession analysis.

| InDel Name | Orientation | Chromosome | Start position (bp) | Primer Sequence (from left to right is 5' to 3') |
|---------------------|-------------|------------|---------------------|--|
| NDSU_IND_1_2.0489 | F | 1 | 2,048,852 | CCAACACTTCGTTCAACAGCCTTTCT |
| NDSU_IND_1_2.0489 | R | 1 | - | ATATGCGCAGCTCCAATTCTCACTTG |
| NDSU_IND_1_4.1957 | F | 1 | 4,195,698 | CACCAGAAAAGAAAAATGAGGTGCAAA |
| NDSU_IND_1_4.1957 | R | 1 | - | CACCAGAAAAGAAAAATGAGGTGCAAA |
| NDSU_IND_1_51.6677 | F | 1 | 51,667,736 | CACAGGGTCATTTGGATCATAGTTCACA |
| NDSU_IND_1_51.6677 | R | 1 | - | CAGCCTATTCTCAGGTGGGTATTCA |
| NDSU_IND_2_22.8464 | F | 2 | 22,846,375 | TTCTCACTTGTTGGGAGTCATCATGC |
| NDSU_IND_2_22.8464 | R | 2 | - | TTCTCACTTGTTGGGAGTCATCATGC |
| NDSU_IND_2_47.6185 | F | 2 | 47,618,518 | TGCAACAATTCGAAATTGGCAGAAA |
| NDSU_IND_2_47.6185 | R | 2 | - | TTGAAACCTTTGACACAACCAAGGAAA |
| NDSU_IND_02_03.5760 | F | 2 | 3,575,961 | GGGTTCGTCCTCTAAATTTGGAGCAGG |
| NDSU_IND_02_03.5760 | R | 2 | - | CCAAAAGTAGGGATAAAGAACAGTGAAGACGG |
| NDSU_IND_3_37.9231 | F | 3 | 37,923,131 | GTTTGCCCCCTGGTGAAGTGGT |
| NDSU_IND_3_37.9231 | R | 3 | - | TCTCTGAATCAACTCCAGCAATAAAAAGGA |
| NDSU_IND_3_48.9580 | F | 3 | 48,957,952 | TGATGTCTTCAGAGCTAAACATCCAGATAGG |
| NDSU_IND_3_48.9580 | R | 3 | - | ATTTTGCGACCAACAGGTGTATGCTT |
| NDSU_IND_03_01.2901 | F | 3 | 1,290,124 | TGTGAGTTAGAGAGGGTTGTTCCGGG |
| NDSU_IND_03_01.2901 | R | 3 | - | TGGATAGAGTGGTTCATGCAAGTACCC |
| NDSU_IND_4_18.2206 | F | 4 | 18,220,611 | GCAAATGCAACATGAGTTGGAAGACA |
| NDSU_IND_4_18.2206 | R | 4 | - | TCAGTTGAAAAGACTGACTGAACACAA |
| NDSU_IND_04_07.4908 | F | 4 | 7,490,819 | TCCATGCTATTAAGGGAGAAGGTCACA |
| NDSU_IND_04_07.4908 | R | 4 | - | TCATTGTCCTTCATCCGTTATCGAATTATCCA |
| NDSU_IND_04_39.7789 | F | 4 | 39,778,927 | GAAGCATAACAGGGAGTGCGAACG |
| NDSU_IND_04_39.7789 | R | 4 | - | ATTGCTCCTGCTCCTGTATCCGG |
| NDSU_IND_5_23.8496 | F | 5 | 23,849,590 | ATCCTGAGTGAGACGAAAGCGACATC |
| NDSU_IND_5_23.8496 | R | 5 | - | GCAAGTCAACTGGCAAGAACAGAACA |

F = forward, R = reverse.

Table A1. InDel Markers used in the intra-accession analysis (continued).

| InDel Name | Orientation | Chromosome | Start position (bp) | Primer Sequence (from left to right is 5' to 3') |
|---------------------|-------------|------------|---------------------|--|
| NDSU_IND_5_40.4088 | F | 5 | 40,408,851 | AAGATGGCCACAAGTGCAATGTCATA |
| NDSU_IND_5_40.4088 | R | 5 | - | CTCCACGCGTCAGGCTATGCTTATTA |
| NDSU_IND_05_02.7232 | F | 5 | 2,723,185 | TCAAAAGTCACTTGCACATAAGCTTGTC |
| NDSU_IND_05_02.7232 | R | 5 | - | TGCTGATTTGTTGGGTTCTTAACATAGCA |
| NDSU_IND_06_15.1145 | F | 6 | 15,114,540 | AGCAAGCATTGGAAAAGTGGGGAG |
| NDSU_IND_06_15.1145 | R | 6 | - | TGAACCACCACCAACACATGCTAC |
| NDSU_IND_06_21.5662 | F | 6 | 21,566,226 | CATAAATCTCTCGTCCACCACCA |
| NDSU_IND_06_21.5662 | R | 6 | - | TGCTGTGGTGAAGATGTTTTGAGTGA |
| NDSU_IND_06_25.5149 | F | 6 | 25,514,894 | TGACGTGATTCCTTGGATCATCAACTCC |
| NDSU_IND_06_25.5149 | R | 6 | - | ACTCATGGCTAAGAGAAAAGTGAACAGTGT |
| NDSU_IND_7_2.5749 | F | 7 | 2,574,873 | ATTACAAGAGTGGATCCGGGTTGACA |
| NDSU_IND_7_2.5749 | R | 7 | - | TGAATTAATTTTATTGCAGAAGGTGGGAAGG |
| NDSU_IND_7_50.8876 | F | 7 | 50,887,583 | TGAGATTTTTAAGGGGAAATGTGCAA |
| NDSU_IND_7_50.8876 | R | 7 | - | CACGTGTCGCATGTGTATAATTTCCAA |
| NDSU_IND_07_35.8798 | F | 7 | 35,879,759 | ACATGCATCACATATCATGCTCTACCAATTCA |
| NDSU_IND_07_35.8798 | R | 7 | - | TGTGAAAGGCTAGACTTACGGACTCTGA |
| NDSU_IND_08_02.0421 | F | 8 | 2,042,109 | GGCAAGAGTGTGTGAAATTGTGGTTGG |
| NDSU_IND_08_02.0421 | R | 8 | - | GCTGTTGATGTTGATGATGTTGCTGCTC |
| NDSU_IND_08_08.6895 | F | 8 | 8,689,541 | AGTCGATATGTGATCTTCAGCACATCCCT |
| NDSU_IND_08_08.6895 | R | 8 | - | TGTTGTCTCACGCACGAGTTGTGA |
| NDSU_IND_08_50.8645 | F | 8 | 50,864,515 | AGTGAATGATGAGGCCGGAGAAGG |
| NDSU_IND_08_50.8645 | R | 8 | - | CTCCATTCCTGTACCTCTTCATCCCTG |
| NDSU_IND_09_35.0297 | F | 9 | 35,029,736 | GCCCGCAAATTGATGATTCCTGGT |
| NDSU_IND_09_35.0297 | R | 9 | - | GATAGCAGCAATTGTTTACTAGAAGCACTGCA |
| NDSU_IND_09_09.0870 | F | 9 | 9,087,048 | TCCGCAAGTTTGATACTCCCATGAGT |
| NDSU_IND_09_09.0870 | R | 9 | - | CCACCGCCCAACCTATCCTACA |
| NDSU_IND_09_17.6034 | F | 9 | 17,603,411 | TGATACAACCTTGTGATCCGTGGTTCAA |

F = forward, R = reverse.

Table A1. InDel Markers used in the intra-accession analysis (continued).

| InDel Name | Orientation | Chromosome | Start position (bp) | Primer Sequence (from left to right is 5' to 3') |
|---------------------|-------------|------------|---------------------|--|
| NDSU_IND_09_17.6034 | R | 9 | - | TGCTACTAGCAATCGTGTTACCAACG |
| NDSU_IND_10_02.1311 | F | 10 | 2,131,117 | TGGGTTTCGCAGTGTATTAAGGTTGTGT |
| NDSU_IND_10_02.1311 | R | 10 | - | TTATGGACGCGCCTTCTTACCTGAG |
| NDSU_IND_10_35.7718 | F | 10 | 35,771,849 | CTGGAGAGAAGTAAGAAAGGGGAATGGTGA |
| NDSU_IND_10_35.7718 | R | 10 | - | CACCCACGCTAAATTCTCCTCCTTTTCT |
| NDSU_IND_10_42.1355 | F | 10 | 42,135,511 | TGTCCTATAATTTATGGACTCGGACGTGTCA |
| NDSU_IND_10_42.1355 | R | 10 | - | CCTGATTGGTCCAAGTGCTCCATTTCT |
| NDSU_IND_11_04.8923 | F | 11 | 4,892,319 | TGTGGTTAACTTGATTTGAGTACACAGCAGA |
| NDSU_IND_11_04.8923 | R | 11 | - | GGATTTTTATCCCAGCACCAACTCACCA |
| NDSU_IND_11_44.9602 | F | 11 | 44,960,210 | GTGATGACGAGTCTTCTTTGTAGATTCTGCA |
| NDSU_IND_11_44.9602 | R | 11 | - | TGGTTAGCATGTGCTTTCTCCTATCACAT |
| NDSU_IND_11_49.2077 | F | 11 | 49,207,665 | ACCTGCACAACACCTGAAGAGACTC |
| NDSU_IND_11_49.2077 | R | 11 | - | TCTGCTGGATCAACCTCTCTCATATCTCA |

F = forward, R = reverse.

Table A2. Significant markers at 0.01 percentile and candidate genes for each trait evaluated by GWAS.

| SNP | Chr | SNP position (bp) | P-value | Bean candidate gene | Marker distance from candidate gene (bp) | Arabidopsis gene symbol | Arabidopsis annotation | R2 for significant peaks |
|--------------------|-----|-------------------|----------|---------------------|--|-------------------------|---|--------------------------|
| Elevation | | | | | | | | |
| Pv02_38216242 | 2 | 38,216,242 | 2.71E-06 | Phvul.002G213600 | 0 | HEN4 | RNA-binding KH domain-containing protein | 0.12 |
| Pv02_38337312 | 2 | 38,337,312 | 4.87E-05 | Phvul.002G215500 | 82,676 | AGL21 | AGAMOUS-like 21 | |
| Pv02_46664311 | 2 | 46,664,311 | 9.30E-06 | Phvul.002G298300 | 0 | - | RING/U-box superfamily protein | 0.09 |
| Pv02_46736576 | 2 | 46,736,576 | 1.32E-05 | Phvul.002G299400 | 2,834 | CBL8 | Calcineurin B-like protein 8 | |
| Pv02_46740440 | 2 | 46,740,440 | 5.64E-06 | Phvul.002G299400 | 6,698 | | | |
| Pv03_39752668 | 3 | 39,752,668 | 1.38E-05 | Phvul.003G175700 | -35,140 | ATDCL4,DCL4 | Dicer-like 4 | 0.08 |
| Pv05_32602302 | 5 | 32,602,302 | 2.36E-05 | Phvul.005G106100 | -2,961 | - | Protein phosphatase 2C family protein | 0.07 |
| Pv05_39204798 | 5 | 39,204,798 | 3.36E-05 | Phvul.005G164300 | 3,241 | - | Leucine-rich repeat protein kinase family protein | 0.07 |
| Pv11_12737204 | 11 | 12,737,204 | 5.33E-06 | NA | NA | NA | NA | |
| Pv11_13054356 | 11 | 13,054,356 | 1.43E-06 | Phvul.011G105600 | 52,664 | CPC | Homeodomain-like superfamily protein | 0.09 |
| Seed length | | | | | | | | |
| Pv02_47021193 | 2 | 47,021,193 | 1.05E-05 | Phvul.002G302000 | 0 | - | SH3 domain-containing protein | |
| Pv02_47294046 | 2 | 47,294,046 | 8.90E-06 | Phvul.002G304600 | 6,346 | - | Pyridoxal phosphate phosphatase-related protein | |
| Pv02_47352798 | 2 | 47,352,798 | 2.74E-05 | Phvul.002G305400 | 0 | - | F-box/RNI-like superfamily protein | 0.11 |
| Pv02_47352799 | 2 | 47,352,799 | 2.87E-05 | Phvul.002G305400 | 0 | | | |
| Pv02_47400959 | 2 | 47,400,959 | 1.75E-05 | NA | NA | NA | NA | |
| Pv02_47556089 | 2 | 47,556,089 | 8.81E-07 | Phvul.002G307600 | 0 | PFN2,PRF2,PRO2 | Profilin 2 | |
| Pv04_46263698 | 4 | 46,263,698 | 2.27E-05 | | | | | 0.05 |

Negative sign in column six indicates that the marker is upstream of the candidate gene and no sign indicates that the marker is downstream of the candidate gene. R-square is calculated for each significant peak, rows in gray facilitate the identification of peaks. Chr = Chromosome. NA = No candidate gene found for the marker.

Table A2. Significant markers at 0.01 percentile and candidate genes for each trait evaluated by GWAS (continued).

| SNP | Chr | SNP position (bp) | P-value | Bean candidate gene | Marker distance from candidate gene (bp) | Arabidopsis gene symbol | Arabidopsis annotation | R2 for significant peaks |
|--------------------------------|-----|-------------------|----------|---------------------|--|-------------------------|---|--------------------------|
| Seed width | | | | | | | | |
| Pv02_47234713 | 2 | 47,234,713 | 2.79E-07 | Phvul.002G304100 | 0 | | | |
| Pv02_47234716 | 2 | 47,234,716 | 2.07E-07 | Phvul.002G304100 | 0 | TET2 | Tetraspanin 2 | |
| Pv02_47234824 | 2 | 47,234,824 | 1.65E-07 | Phvul.002G304100 | 0 | | | |
| Pv02_47262606 | 2 | 47,262,606 | 4.44E-08 | Phvul.002G304500 | 13,630 | - | Protein kinase protein with adenine nucleotide alpha hydrolases-like domain | 0.11 |
| Pv02_47273758 | 2 | 47,273,758 | 5.80E-07 | Phvul.002G304500 | 2,478 | | Phosphotyrosine protein phosphatases superfamily protein | |
| Pv02_47516312 | 2 | 47,516,312 | 6.13E-07 | Phvul.002G307100 | 8,159 | - | | |
| Seed width/length ratio | | | | | | | | |
| Pv02_47234716 | 2 | 47,234,716 | 1.01E-10 | Phvul.002G304100 | 0 | TET2 | Tetraspanin 2 | |
| Pv02_47234824 | 2 | 47,234,824 | 3.48E-11 | Phvul.002G304100 | 0 | | | |
| Pv02_47262606 | 2 | 47,262,606 | 3.45E-11 | Phvul.002G304500 | 13,630 | - | Protein kinase protein with adenine nucleotide alpha hydrolases-like domain | 0.25 |
| Pv02_47400959 | 2 | 47,400,959 | 4.56E-11 | NA | NA | NA | NA | |
| Pv02_47556089 | 2 | 47,556,089 | 1.23E-11 | Phvul.002G307600 | 0 | PFN2,PRF2,PRO2 | Profilin 2 | |
| Seed coat luster | | | | | | | | |
| Pv02_23321712 | 2 | 23,321,712 | 3.97E-06 | Phvul.002G108800 | 47,394 | - | Chalcone-flavanone isomerase family protein | 0.06 |
| Pv02_23604335 | 2 | 23,604,335 | 1.20E-06 | NA | NA | NA | NA | |
| Pv02_23604337 | 2 | 23,604,337 | 1.40E-06 | NA | NA | NA | NA | |
| Pv07_17581536 | 7 | 17,581,536 | 8.90E-06 | Phvul.007G119178 | 87,993 | | | |
| Pv07_17581555 | 7 | 17,581,555 | 7.23E-06 | Phvul.007G119178 | 88,012 | | Bifunctional inhibitor/lipid-transfer protein/seed storage | 0.08 |
| Pv07_17581604 | 7 | 17,581,604 | 7.23E-06 | Phvul.007G119178 | 88,061 | - | 2S albumin superfamily protein | |
| Pv07_17601477 | 7 | 17,601,477 | 4.13E-06 | Phvul.007G119178 | 107,934 | | | |

No sign in column six indicates that the marker is downstream of the candidate gene R-square is calculated for each significant peak, rows in gray facilitate the identification of peaks. Chr = Chromosome. NA = No candidate gene found for the marker.

Table A2. Significant markers at 0.01 percentile and candidate genes for each trait evaluated by GWAS (continued).

| SNP | Chr | SNP position (bp) | P-value | Bean candidate gene | Marker distance from candidate gene (bp) | Arabidopsis gene symbol | Arabidopsis annotation | R2 for significant peaks |
|------------------------|-----|-------------------|----------|---------------------|--|-------------------------|--|--------------------------|
| Flower color | | | | | | | | |
| Pv06_2490072 | 6 | 2,490,072 | 8.75E-06 | NA | NA | NA | NA | 0.07 |
| Pv06_2502906 | 6 | 2,502,906 | 1.83E-05 | NA | NA | NA | NA | |
| Pv06_7000786 | 6 | 7,000,786 | 5.06E-06 | Phvul.006G015400 | -46,614 | | | 0.09 |
| Pv06_7001628 | 6 | 7,001,628 | 9.12E-08 | Phvul.006G015400 | -45,772 | CYP75B1, D501,TT7 | Cytochrome P450 superfamily protein | |
| Pv06_7001680 | 6 | 7,001,680 | 9.83E-07 | Phvul.006G015400 | -45,720 | | | |
| Pv06_7006012 | 6 | 7,006,012 | 1.06E-05 | Phvul.006G015400 | -41,388 | | | |
| Pv06_11349339 | 6 | 11,349,339 | 1.78E-06 | NA | NA | NA | NA | 0.07 |
| Cotyledon color | | | | | | | | |
| Pv02_9064264 | 2 | 9,064,264 | 1.38E-04 | Phvul.002G065600 | -20,144 | ATHIR1,HIR1 | SPFH/Band 7/PHB domain-containing membrane-associated protein family | 0.05 |
| Pv03_5281445 | 3 | 5,281,445 | 1.53E-04 | NA | NA | NA | NA | 0.03 |
| Pv04_6146304 | 4 | 6,146,304 | 7.59E-05 | NA | NA | NA | NA | 0.04 |
| Pv05_10309053 | 5 | 10,309,053 | 1.95E-04 | NA | NA | NA | NA | 0.04 |
| Pv06_7001628 | 6 | 7,001,628 | 1.24E-05 | Phvul.006G015400 | -45,772 | CYP75B1, D501,TT7 | Cytochrome P450 superfamily protein | 0.06 |
| Pv06_7001680 | 6 | 7,001,680 | 2.37E-05 | Phvul.006G015400 | -45,720 | | | |
| Pv08_3852507 | 8 | 3,852,507 | 1.73E-04 | NA | NA | NA | NA | 0.04 |
| Stem color | | | | | | | | |
| Pv06_2490072 | 6 | 2,490,072 | 1.01E-06 | NA | NA | NA | NA | 0.09 |
| Pv06_2502906 | 6 | 2,502,906 | 1.66E-06 | NA | NA | NA | NA | |
| Pv06_5163863 | 6 | 5,163,863 | 2.58E-05 | NA | NA | NA | NA | 0.08 |
| Pv06_5253572 | 6 | 5,253,572 | 1.15E-05 | NA | NA | NA | NA | |
| Pv06_7001628 | 6 | 7,001,628 | 1.33E-05 | Phvul.006G015400 | -45,772 | CYP75B1, D501,TT7 | Cytochrome P450 superfamily protein | 0.06 |
| Pv06_7001680 | 6 | 7,001,680 | 2.47E-05 | Phvul.006G015400 | -45,720 | | | |

Negative sign in column six indicates that the marker is upstream of the candidate gene. Chr = Chromosome. NA = No candidate gene found for the marker.

Table A2. Significant markers at 0.01 percentile and candidate genes for each trait evaluated by GWAS (continued).

| SNP | Chr | SNP position (bp) | P-value | Bean candidate gene | Marker distance from candidate gene (bp) | Arabidopsis gene symbol | Arabidopsis annotation | R2 for significant peaks |
|-------------------------------|-----|-------------------|----------|---------------------|--|-------------------------|---|--------------------------|
| Pod color | | | | | | | | |
| Pv08_3153417 | 8 | 3,153,417 | 7.30E-06 | Phvul.008G038000 | 10,411 | | | |
| Pv08_3153521 | 8 | 3,153,521 | 7.30E-06 | Phvul.008G038000 | 10,515 | ATMIXTA,ATMYB16,MYB16 | myb domain protein 16 | 0.06 |
| Pv08_3153528 | 8 | 3,153,528 | 7.30E-06 | Phvul.008G038000 | 10,522 | | | |
| Pv08_3153534 | 8 | 3,153,534 | 7.30E-06 | Phvul.008G038000 | 10,528 | | | |
| Rust resistance | | | | | | | | |
| Pv02_42867003 | 2 | 42,867,003 | 1.05E-04 | Phvul.002G256600 | -631 | ATUBP3,UBP3 | Ubiquitin-specific protease 3 | |
| Pv02_42893682 | 2 | 42,893,682 | 9.07E-05 | Phvul.002G257000 | -38,435 | | Leucine-rich repeat transmembrane protein kinase family protein | 0.06 |
| Pv02_42933517 | 2 | 42,933,517 | 5.16E-05 | Phvul.002G257000 | 0 | - | | |
| Pv02_42990094 | 2 | 42,990,094 | 1.02E-04 | Phvul.002G257800 | 5,414 | - | RPM1-interacting protein 4 (RIN4) family protein | |
| Pv08_2277889 | 8 | 2,277,889 | 1.09E-04 | Phvul.008G028200 | 0 | | Malectin/receptor-like protein kinase family protein | 0.04 |
| Pv08_2303436 | 8 | 2,303,436 | 4.97E-05 | Phvul.008G028600 | 1,202 | FER | | |
| Anthracnose resistance | | | | | | | | |
| Pv01_3734105 | 1 | 3,734,105 | 3.26E-04 | Phvul.001G044900 | 0 | | Protein kinase superfamily protein | |
| Pv01_3734112 | 1 | 3,734,112 | 1.87E-04 | Phvul.001G044900 | 0 | - | | 0.09 |
| Pv01_4836007 | 1 | 4,836,007 | 1.32E-04 | Phvul.001G046400 | 16,048 | - | Glycosyl hydrolase family protein with chitinase insertion domain | |
| Pv01_6220257 | 1 | 6,220,257 | 6.15E-05 | NA | NA | NA | NA | |
| Pv01_30004721 | 1 | 30,004,721 | 2.92E-04 | NA | NA | NA | NA | 0.05 |
| Pv03_33052812 | 3 | 33,052,812 | 3.15E-04 | Phvul.003G134100 | -57,787 | ATMKK2,MK1, MKK2 | MAP kinase kinase 2 | 0.05 |
| Pv05_30232782 | 5 | 30,232,782 | 2.38E-04 | Phvul.005G097400 | 12,350 | - | GTP-binding protein, HflX | 0.05 |

Negative sign in column six indicates that the marker is upstream of the candidate gene and no sign indicates that the marker is downstream of the candidate gene. R-square is calculated for each significant peak, rows in gray facilitate the identification of peaks. Chr = Chromosome. NA = No candidate gene found for the marker.

Table A 2. Significant markers at 0.01 percentile and candidate genes for each trait evaluated by GWAS (continued).

| SNP | Chr | SNP position (bp) | P-value | Bean candidate gene | Marker distance from candidate gene (bp) | Arabidopsis gene symbol | Arabidopsis annotation | R2 for significant peaks |
|--------------------------------|-----|-------------------|----------|---------------------|--|-------------------------|-------------------------------------|--------------------------|
| Phytophthora resistance | | | | | | | | |
| Pv04_2305784 | 4 | 2,305,784 | 1.92E-05 | Phvul.004G019600 | -12,324 | | | |
| Pv04_2305787 | 4 | 2,305,787 | 1.92E-05 | Phvul.004G019600 | -12,321 | ATCSLC04,CSLC04 | Cellulose-synthase-like C4 | 0.05 |
| Pv06_21241445 | 6 | 21,241,445 | 2.89E-06 | Phvul.006G102300 | -28,172 | HEL,PR-4,PR4 | Pathogenesis-related 4 | |
| Pv06_26759529 | 6 | 26,759,529 | 3.77E-06 | Phvul.006G164700 | -5,137 | | Protein of unknown function, DUF538 | |
| Pv06_26760100 | 6 | 26,760,100 | 2.83E-06 | Phvul.006G164700 | -4,566 | - | | 0.13 |
| Pv06_26845459 | 6 | 26,845,459 | 7.27E-06 | Phvul.006G165300 | 0 | | | |
| Pv06_26853505 | 6 | 26,853,505 | 7.99E-09 | Phvul.006G165300 | 0 | ATCUL4,CUL4 | Cullin4 | |

Negative sign in column six indicates that the marker is upstream of the candidate gene and no sign indicates that the marker is downstream of the candidate gene. R-square is calculated for each significant peak, rows in gray facilitate the identification of peaks. Chr = Chromosome. NA = No candidate gene found for the marker.

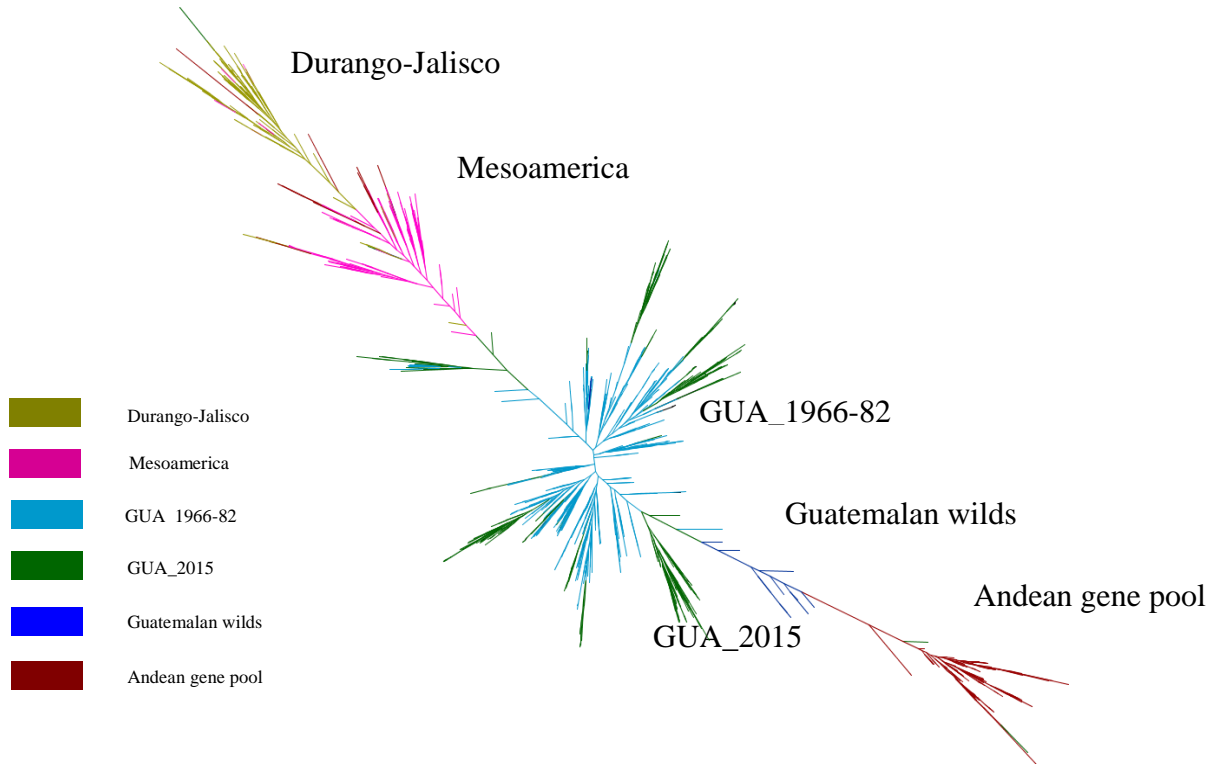


Figure A1. Results of the maximum likelihood tree including all the populations analyzed.

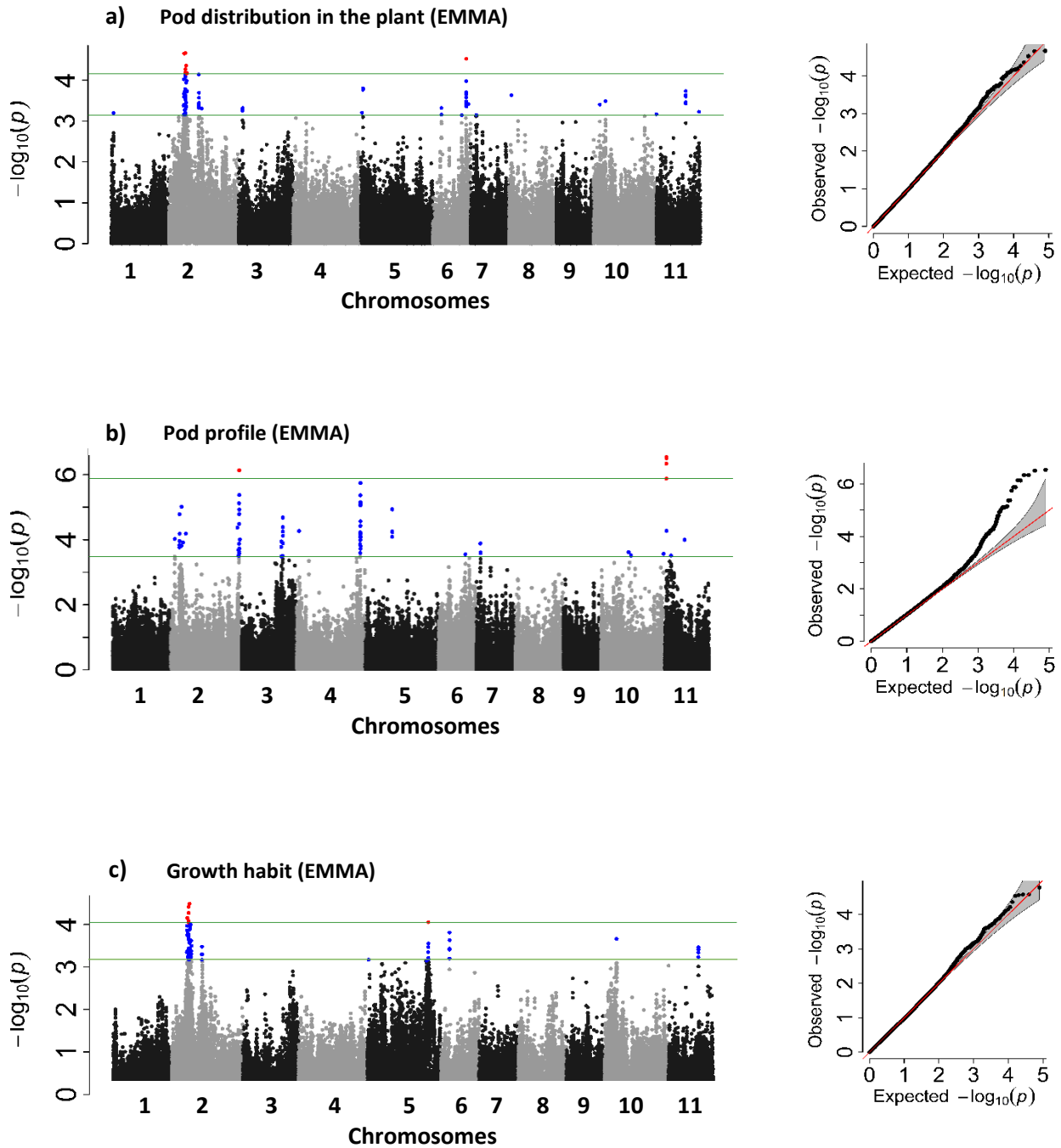


Figure A2. Manhattan plots and QQ plots for the best models of the eight agronomic traits that were not analyzed in the discussion. a) Pod distribution in the plant; b) Pod profile; c) Growth habit. Green lines represent the cutoff values for 0.1 and 0.01 percentiles. Markers significant for the 0.01 and 0.1 percentiles are colored in red and blue, respectively. Best model is indicated in parenthesis.

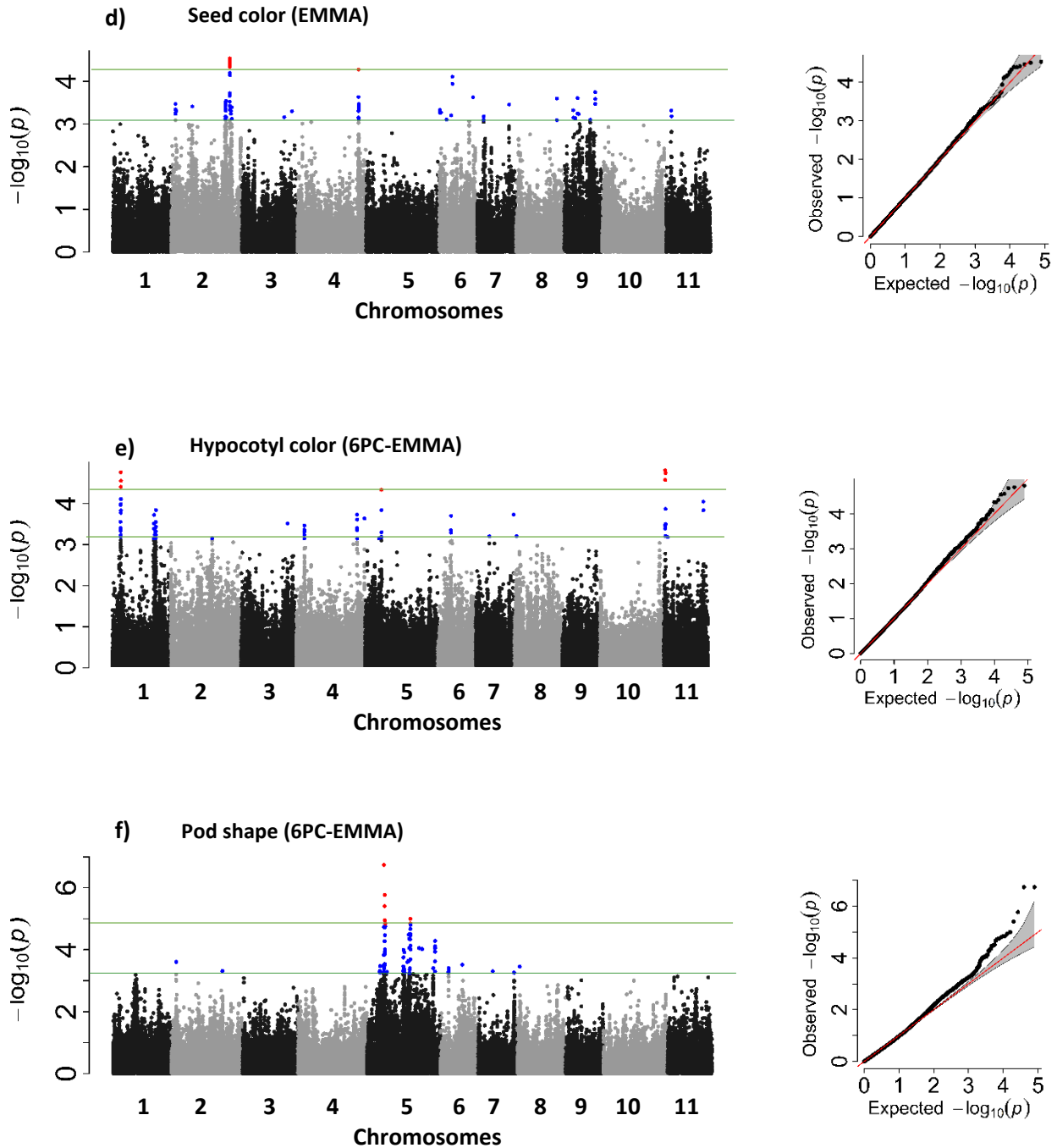


Figure A2. Manhattan plots and QQ plots for the best models of the eight agronomic traits that were not analyzed in the discussion (continued). d) Seed color; e) Hypocotyl color; f) Pod shape. Green lines represent the cutoff values for 0.1 and 0.01 percentiles. Markers significant for the 0.01 and 0.1 percentiles are colored in red and blue, respectively. Best model is indicated in parenthesis.

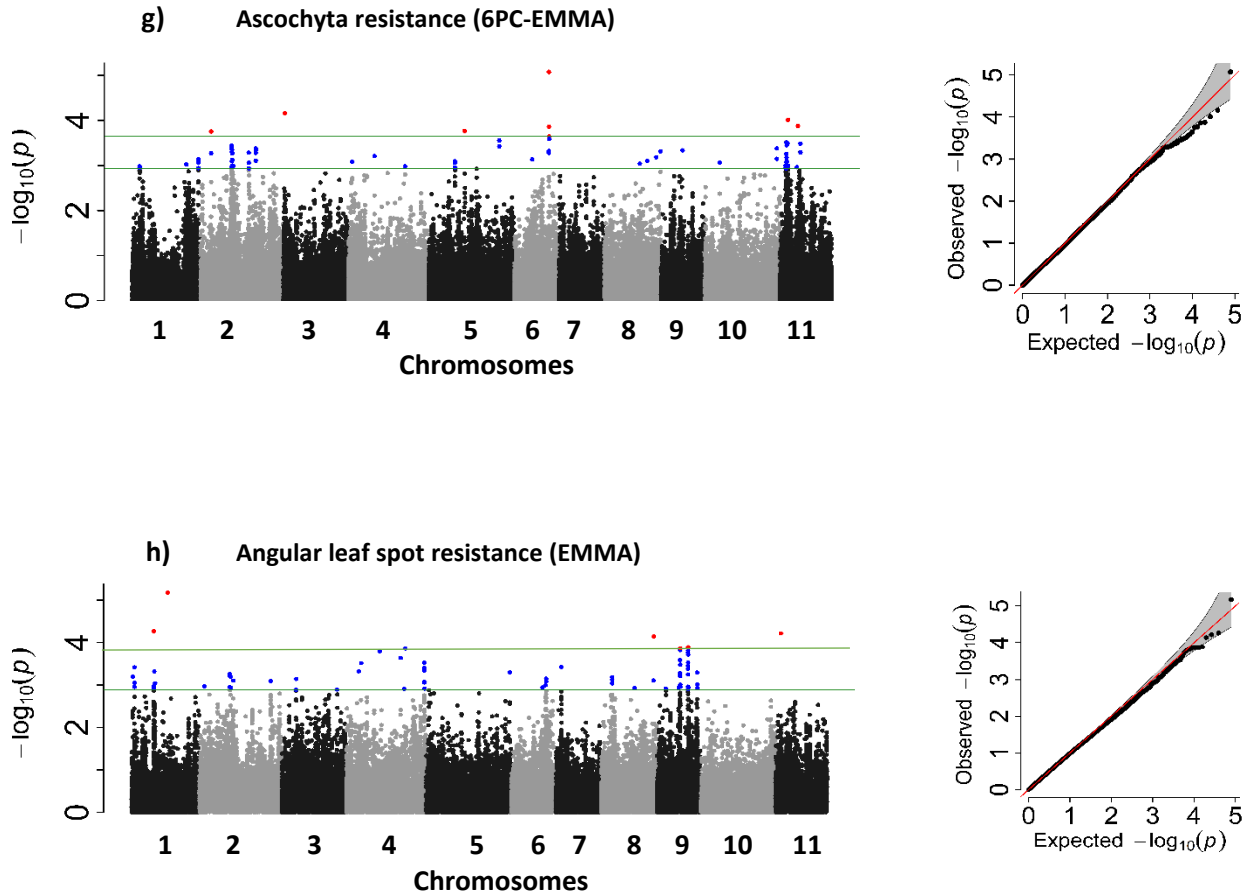


Figure A2. Manhattan plots and QQ plots for the best models of the eight agronomic traits that were not analyzed in the discussion (continued). g) Ascochyta resistance; h) Angular leaf spot resistance. Green lines represent the cutoff values for 0.1 and 0.01 percentiles. Markers significant for the 0.01 and 0.1 percentiles are colored in red and blue, respectively. Best model is indicated in parenthesis.