ABSTRACT

VILLEGAS CHIRINOS, FRANCO. Breeding for Early Maturity in Peanuts (*Arachis hypogaea* L.) using Traditional Methods and Marker Assisted Selection (MAS). (Under the direction of Drs. Thomas G. Isleib and Susana R. Milla-Lewis).

Peanut (*Arachis hypogaea* L.) is an economically important legume grown in world-wide. In the U.S.A. is marketed mainly as whole seed as a snack or for the manufacture of processed products such as peanut butter and candy. Early maturity is an urgently needed trait in regions with short growing seasons such as North Carolina. The main objectives of this investigation were (1) to study the inheritance of early maturity, and (2) to map QTL associated with this trait. Two populations of RILs were used for these purposes: Pop1 (132 RILs) with PI 313949 (late, large-seeded) and Chico-ol_1ol_2 (early, small-seeded) as parents, and Pop2 (50 RILs) with PI 365550 (late, large-seeded) and Chico-ol_1ol_2 as parents. In addition to maturity index, yield (Kg plot^{-1}), traits related with fruit and seed size (length, width, weight), number of seeds per pod, and fruit color (Hunter L, a, and b scores) were measured for two years with two reps per year, and variance components for each trait were obtained. Additive variance (σ_A^2) and additive-by-additive epistatic variance (σ_{AA}^2) were estimated. Estimates of σ_A^2 for maturity index, width of 20 fruits, and width of 20 seeds were negative in both populations. Estimates of σ_A^2 for number of seeds in 20 fruits and pod redness were negative only in Pop1. Estimates of σ_A^2 for length and weight of 20 fruits, pod redness, and pod yellowness were negative only in Pop2. Estimates of σ_{AA}^2 for all traits in both populations were positive except for number of seeds in 20 fruits and pod redness in

Pop2. These results suggest that epistatic gene action operates in the populations tested and that the effects of individual alleles have little or no influence on the observed phenotypes. Pop1, Pop2, and the combined information for both populations (182 RILs total) were analyzed and produced three similar linkage maps with slightly variable numbers of loci and LG (with more information producing larger maps with more LG). Three QTL associated with early maturity were identified in these populations. QTL-Maturity I was significant in Pop1 (LOD=4.2 and R^2=0.13) and in the combined population (LOD=3.13 and R^2=0.07). QTL-Maturity II was significant only in Pop2 (LOD=2.52 and R^2=0.34), but regions with high LOD scores that did not reach the significance threshold were found in corresponding regions in Pop1 and in the combined population. QTL-Maturity III was significant only in the combined population (LOD=3.69 and R^2=0.10). MAS breeding for early maturity seems feasible for maturity index in this or related populations, with QTL-Maturity I and its associated markers, GM34 and GM2689, as the most promising finding for that purpose because of its higher significance, usable R^2 value, and consistent location in two out of three populations. The combined results of this investigation encourage the development of a MAS breeding program to improve maturity in peanuts.

Breeding for Early Maturity in Peanuts (*Arachis hypogaea* L.) using Traditional Methods and Marker Assisted Selection (MAS)

by
Franco Villegas Chirinos

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Crop Science

Raleigh, North Carolina

2011

APPROVED BY:

Dr. Thomas G. Isleib Committee Chair	Dr. Susana R. Milla-Lewis Committee Co-chair
Dr. David L. Jordan	Dr. Consuelo Arellano

DEDICATION

First, I would like to dedicate this work to God, without Him nothing would be possible. Secondly, to the persons who make me feel the love of God in the most direct way possible, my beloved family: Mamá, Papá, Julio, Lula, Miguel, and Nala. Also, to my dear bosses and friends, Susana and Tom. Finally, to all the wonderful people who touched my life during this process.

BIOGRAPHY

Franco Villegas Chirinos was born in Lima, Perú, in 1982. He grew up in a middle class family with professional parents. In the early stages of his life he was constantly mentally challenged by his father, Jorge Franco Villegas Clavo, the most brilliant mind he has ever met. Because of this early stimulation and increasing curiosity, Franco decided to study sciences, with emphasis in Biology, at Universidad Agraria La Molina (UNALM) in Lima. He got his bachelor in science degree at UNALM in 2005. He did an internship in the Tropical Medicine Institute of Universidad Cayetano Heredia in 2005, worked in the Mycology and Biotechnology Laboratory (LMB) of UNALM in 2006, and worked as a professional trainee in the Virology laboratory of the International Potato Center (CIP) in 2007. In 2008, he emigrated to the U.S.A. to pursue graduate studies at North Carolina State University (NCSU). Through his time at NCSU he became acquainted with traditional breeding topics such as variance components and heritability estimates, and also to modern technologies such as QTL mapping to help in peanut breeding efforts. Thanks to the exposure to this wide array of approaches (and a research that resulted controversial sometimes), Franco, is well prepared to continue his career. He looks forward to starting a PhD, with the ultimate goal of becoming a good scientist as a plant breeder.

ACKNOWLEDGMENTS

First, I would like to thank my advisers, Dr. Susana R. Milla-Lewis and Dr. Thomas G. Isleib. I have a huge debt with these scientists for their patience, support, and help. From the application process to grad school to the final reviews of my thesis, they have been a huge relief in this difficult and sometimes painful process.

I would also like to thank all the peanut breeding program team at NCSU: Susan Copeland, Phillip Rice, Hugo Pinto, Edward Hassel, Gaines Bailey, Katie Kroening**,** Haley Mann , Meredith Holder, and all the people that passed through the program during my time there. Your help and friendship made this work possible.

I would also like to thank the turf breeding team at NCSU: Carolina Zuleta, Jennifer Kimball, Bessy Raudales, Steven Mulkey, and Tito Claure. We shared a lot of good times in our cramped graduate student office in Williams Hall. Thanks for your friendship, support and patience, especially when I used all my capabilities to try to distract you from work.

Finally, I would like to thank my family, all my cousins, aunts, and uncles (Villegas and Chirinos), my best friends: Gustavo Zapata, Luis Ernesto Tohara, Mario Torres, and Enrique Seminario. Their unconditional support and love helped me in times of huge necessity.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER III. Use of Simple Sequence Repeat (SSR) Markers for Mapping Quantitative Trait Loci (QTL) Influencing Early Maturity in *Arachis hypogaea* L.

Chapter I

Literature Review

Arachis hypogaea L. (Peanut)

Taxonomy and market types. Peanut (*Arachis hypogaea* L.) is classified in subtribe Stylosanthinae of tribe Aeschynomeneae of family Leguminosae (Hammons, 1982). The genus *Arachis* is subdivided in nine sections according to morphological traits, ploidy level, and chromosome number (cytogenetic characteristics). These sections are Arachis, Caulorrhizae, Erectoides, Extranervosae, Heteranthae, Procumbentes, Rhizomatosae, Trierectoides, and Triseminate. *A. hypogaea* is the type species in section Arachis. It has tetrafoliolate leaves, *i.e.*, having four leaflets per leaf, with decumbent or erect plants, and pegs penetrate the soil at an angle close to 45 degrees. This section includes diploid (2n=2x=20) and tetraploid (2n=4x=40) species.

Due to its economic importance, *A. hypogaea* is the most studied and described species of the genus (Moss and Rao, 1995). A large amount of variability in growth habits and morphological traits, often with agronomic importance, is found among members of the cultivated peanut. This variability has led to the classification of peanuts into several sub-specific taxa, two subspecies (subsp. *hypogaea* and subsp. *fastigiata* Waldron) and six botanical varieties, vars. *hypogaea* and *hirsuta* Köhler in subsp. *hypogaea* and vars. *fastigiata*, *vulgaris* Harz, *peruviana* Krapov. & Gregory, and *aequatoriana* Krapov. & Gregory in subsp. *fastigiata* (Krapovickas and Gregory, 1994).

A somewhat parallel classification based on peanut pod and seed dimensions exists in the U.S. market. There are four U.S. market types of cultivated peanuts: virginia, runner, spanish and valencia. The virginia market type accounts for 18% of U.S. peanut production. Peanuts of this market type have larger pods (more than 2000 vs. less than 1800 mg pod^{-1})

and seeds (more than 800 vs. less than 700 mg seed^{-1}) than those of the runner market type (Knauft *et al.*, 1987). Virginia-type peanuts are sold in the in-shell market or can be used as salted and cocktail peanuts after shelling. In contrast, the runner market-type, which accounts for roughly 80% of U.S. peanut production, is used for the manufacture of peanut butter, candy and peanut oil. Both the virginia and runner market types are genetically based primarily on ancestry from var. *hypogaea* with some admixture of fastigiate ancestry (Isleib and Wynne, 1992; Isleib *et al.*, 2001). The spanish market type occupies approximately 5% of U.S. peanut acreage, especially in the Southwestern production area of Texas and Oklahoma. Spanish peanuts have uses similar to the runner market type and are almost exclusively based on var. *vulgaris* ancestry. The valencia market-type corresponds to var. *fastigiata* and is produced in small proportion in relation to the total U.S. peanut production (approximately 1%). Valencia-type peanuts are consumed roasted in the shell or boiled (Knauft *et al.*, 1987; USDA NASS, 2011a).

Economic importance. *A. hypogaea* is grown extensively in tropical, subtropical, and warm temperate regions of the world. After soybean (*Glycine max* [L.] Merr.), rapeseed (*Brassica napus* L.) and cottonseed (*Gossypium hirsutum* L.), peanut is the fourth most common crop grown for oilseed production globally (USDA FAS, 2011a). However, the uncrushed whole seed is edible and highly appreciated as human food and in processed products. This is especially so in the U.S. where peanut butter, peanut snacks and candy are the main forms in which peanut is consumed. Peanut derivatives, such as flour, peanut meal, and protein isolates, are useful as protein supplements in the production of fortified foods that can aid in nutrition; especially in some developing countries where animal protein is too

expensive for the common person (Singh and Singh, 1991). In addition to the value of the seeds as human food and oil stock, the vegetative residue of the crop can be used to feed cattle. In some African and Asian countries, the use of peanut stover (or "haulms") is most economically important in times of shortage of fodder (Moss and Rao, 1995).

The projected global production of peanuts for 2011 is 33 million metric tons (mmt). In this projection, China is the leading producer (14.7 mmt), followed by India (4.9 mmt), the U.S.A. with 1.68 mmt, and Nigeria with 1.55 mmt. It is noticeable that the U.S.A. has reduced its peanut production by 28.2% from the previous year (USDA FAS, 2011b). In the U.S.A., North Carolina was the fifth state in peanut production for 2010 with approximately 6% of the total value of production for the country. Georgia was the leading state with 44% of production (USDA NASS, 2011b). The value of the North Carolina peanut crop was 57.8 million dollars in 2010; 401 million dollars in Georgia, and the total for the country was 901 million dollars.

Cytogenetics and origins. *A. hypogaea* is an allotetraploid (2n=4x=40) with two sets of homoeologous chromosomes, named the "A" and "B" genomes, resulting in *A. hypogaea*'s designation as an "AABB" species. Chromosomes of the "A" genome are larger and "B" genome chromosomes show a characteristic secondary constriction (Stalker *et al.*, 1995). *A. hypogaea* tends to behave as a diploidized tetraploid during meiosis. However, low rates of occurrence of quadrivalents evidencing the formation of chiasmata between the four sets of chromosomes have been reported (Stalker *et al.*, 1995). The formation of the peanut genome involved the union of two closely related but different diploid genomes, followed by a rare event of polyploidization (chromosome duplication) and reproductive

isolation of perhaps one or a few natural interspecific hybrids (Moretzsohn *et al.*, 2005). These tetraploid hybrids were unable to produce fertile offspring with surrounding diploid (2n=2x=20) *Arachis* species because of the difference in chromosome number (Singh, 1985; Singh and Gibbons, 1986). The result of this sexual isolation was a disruption of gene flow from the few reproductively isolated mutant plants that probably gave rise to all the cultivated peanut landraces (Hopkins *et al.*, 1999). The natural variation observed within *A. hypogaea* resulted from mutation, recombination, and selection that occurred after the polyploidization event, a short period compared with the age of the highly variable diploid species.

Quantitative Traits

Most of the variability found in nature, including many of the traits important for agriculture, shows continuous phenotypic variation without clear-cut categories (Falconer and Mackay, 1996). When evaluating these characters, one cannot classify or separate individuals into discrete groups without defining arbitrary threshold values. For example, when evaluating height among a population of plants one cannot define groups, *i.e.*, "tall" and "short", unless one arbitrarily chooses a critical value to separate the groups. However, the formed classes do not represent the continuum in height observed among individuals. There are two main reasons why this continuous variation arises. First, the additive action of many polymorphic genes, each influencing a small fraction of the final phenotype. Second, the confounding effect of truly continuous variation originating from the environment, genotype-by-environment interaction, and non-additive effects of the genes involved

(dominance, epistasis, and higher-order interactions) in the trait (Falconer and Mackay, 1996). The relative importance of each of these sources of variability and their interaction depends on the specific trait and biological system under evaluation. One must bear in mind that the distinction between genes governing qualitative Mendelian characters (for which clearly cut classes are observable) and those concerned with quantitative traits lies in the magnitude of their effects relative to other sources of variation. A gene with an effect large enough to cause a recognizable discontinuity even in the presence of segregation at other loci and of non-genetic variation can be studied as a gene that follows Mendelian laws, whereas a gene whose effect is not large enough to cause a discontinuity cannot be studied individually (Falconer and Mackay, 1996). In agriculture, there are many quantitative traits of critical importance such as yield, maturation time, plant size, flowering time, etc. (Holland, 2007). These traits are difficult to improve using traditional breeding methods because the effects of individual genes affecting the final phenotype are not easily recognizable (Dudley, 1993).

Early maturity. Early or short-duration genotype is a relative term. Depending on latitude and weather patterns, a peanut variety could be classified as early-maturing with times from sowing to harvest of 140 days, in some regions of the U.S.A., to less than 90 days in some regions of West Africa with short rainy seasons (Nigam and Aruna, 2008). Peanut maturity involves complex biochemical processes in the developing kernel such as stabilization of the protein content in the last days of maturation or carbohydrate content diminution accompanied by oil content increment, both occurring throughout the process (Rowland *et al.*, 2006). Maturity is a quantitative trait influenced by many genes and the environment. The highly indeterminate fruiting pattern of the peanut plant, and the fact that

peanut pods grow underground covered from view, make the prediction of peanut maturity a difficult task with potentially large economic consequences if an incorrect decision is made (Pattee *et al.*, 1974). A tendency of mature peanuts to develop more intense desired flavors and less intense "off" (undesired) flavors after roasting has been reported (Sanders *et al*., 1989). On the other hand, peanut lots containing high percentages of immature pods have lower potential for development of high intensities of the critical roasted peanut sensory attribute and higher potential for short shelf life due to intensified "off" flavors (Sanders *et al.*, 1989).

In order to address the issue of accurate determination of peanut maturity, several methods have been developed. These methods can be separated into four groups based on the feature evaluated (Sanders *et al.*, 1982a). One group includes indirect methods such as counting days after planting (DAP) and the heat unit system. The latter method calculates the necessary time from planting to optimum digging date using a formula based on daily maximum and minimum temperatures (Rowland *et al.*, 2006). A second group includes techniques that evaluate some relative color such as the internal hull color, oil color, methanol extract, the shell-out percentage, and the pod maturity profile. A third group uses weight and/or weight relationships to assess maturity measuring kernel weight, kernel density, or the seed hull weight ratio (seed/hull ratio maturity index, SHMI). Finally, a fourth group includes methods based in the quantitation of a specific component whose concentration change is correlated with maturity stage. This group includes the arginine maturity index (AMI) and the arachin polypeptide determination (Sanders *et al.*, 1982a; Grimm *et al.*, 1998). The currently preferred method to assess maturity in peanuts is the pod

maturity profile or hull-scrape method developed by William and Drexler (1981) (Grimm *et al.*, 1998; Sanders *et al.*, 1982a; Sanders et al 1982b). Using this method the peanut fruit exocarp is removed by abrasion. Subsequently, fruits are separated into categories reflecting differential maturity based on mesocarp color. These categories are: white, yellow, orange, brown or black. Finally, the number of black and brown fruits (mature pods) over the total number of blasted fruits is used as an estimator of maturity called maturity index.

Breeding early maturing cultivars with improved yield and good agronomic characteristics is a major objective of many peanut programs. This is especially true for breeders in regions of the world with short growing seasons characterized by end-of-season droughts, or ones with cooler temperatures and early frosts (Upadhyaya *et al.*, 2006; Bell *et al.*, 1994). Furthermore, early cultivars with short growing seasons would be better for intercropping systems because they offer less competition to later maturing crops and would be able to escape late season diseases and insect pests (Nigam and Aruna, 2008). Shortened growing seasons occur in the Virginia-Carolina and west Texas peanut growing regions where cool night temperatures late in the growing season can retard maturation and cause incomplete seed filling. It must be noted that immature pods have an important negative economic impact for growers because if present in high proportion among harvested pods, they lower peanut yield and grade or quality (Sanders, 1989). Furthermore, immature peanuts are more prone to contamination by toxigenic fungi of the genus *Aspergillus* that produce aflatoxins (toxic and carcinogenic compounds) (Upadhyaya, 2006) and to the development of the undesirable “fruity / fermented” sensory attribute (Sanders *et al.*, 1989).

Pod color and brightness. These quantitative traits have become increasingly important in the last decade, especially for peanut shellers in the Virginia-Carolina region. For the in-shell peanut market, the brightness and color of the pod influence the consumer's decision to buy a particular brand of peanuts, *i.e.*, U.S. consumers are averse to purchasing peanuts with dark pods. For this reason, breeders of peanuts destined for the in-shell market must pay attention to these traits and incorporate measurements of visual aesthetics into their grading procedures prior to selecting the best genotypes (Isleib *et al.*, 1997).

Objective measurements of colors in the visible region can be obtained by various electronic devices, among them the Hunterlab colorimeter (Hunter Associates Laboratory, Inc., Reston, VA) has proved to be practical and efficient (Isleib *et al.*, 1997; Grimm *et al.*, 1998). In the Hunter color system a tridimensional color space is defined by three variables: Hunter L, Hunter a, and Hunter b scores (Grimm *et al.*, 1998). This instrument illuminates the sample and detects reflected light with three different sensors, delivering three different values: the Hunter L score is a measure of the brightness or albedo ratio of reflected to incident light of the sample with a score of 0 indicating complete blackness and 100 indicating perfect whiteness, the Hunter a score assesses color in a red-green scale with increasingly positive scores indicating more intense red color and negative scores signifying green color, and the Hunter b score assesses color on a blue-yellow scale with positive values of increasing magnitude for increasing intensities of yellow color and negative values for different intensities of blue colors (Isleib *et al.*, 1997).

Yield, pod and seed dimensions. These quantitative traits are of extreme importance for peanut growers. Yield is one of the traits to be considered in any peanut breeding

program before the release of a new cultivar, for the obvious reason that it influences farmers' revenue. Pod and seed dimensions define the four U.S. market types (virginia, spanish, valencia, and runner). Peanut pod grades, determined using screens with different opening sizes, ultimately determine the prices that the farmers receive for their lots. Consequently, breeding for lines with the correct size and shape, relative to the target market-type, is an important goal of peanut breeding programs (Davidson *et al.*, 1982).

Heritability and variance components

Heritability and analysis of sources of variability are very important for the efficacy of plant breeding methods. Using heritability estimates a breeder is able to assess and compare alternative selection strategies, in terms of the expected genetic gains. For example, if the desired trait exhibits low heritability, it would require evaluation in several different environments (locations and/or years) to get a good estimate of the true genotypic value and perform efficient selection. Meanwhile, a trait with high heritability would require a minimal amount of effort and resources (probably only one or two environments) to determine the "real" value for the trait and make selections that will be useful for the breeder's purposes, *i.e.*, to modify the allelic frequencies.

A useful tool in breeding that requires understanding of heritability is indirect selection. Indirect selection in a broad sense is when the selected plants are not used directly. On the contrary, selection is based on the correlation between the selected plants and the plants that are actually been used later (Wricke and Weber, 1986). However, indirect selection usually refers to selecting for a trait with high heritability in order to improve a correlated trait with

low heritability. Indirect selection is advantageous in many different situations, for example when the desired trait is expressed late in the life cycle of the organism, *i.e.* using plantlet height to estimate tree yield, or when a trait is greatly influenced by the environment, *i.e.* using a correlated trait to estimate yield. Indirect selection is the basis of MAS, because we are using a "trait" with perfect heritability to select for traits that are difficult to measure or have very low heritability as is the case of maturity in peanuts. Molecular marker information is, in theory, perfectly heritable because the DNA sequence is assessed directly. However, we must keep in mind that indirect selection in any case is only more advantageous than direct selection if the indirect character can be measured more accurately than the direct trait. Furthermore, it only makes sense to use MAS if the correlation between the predicted phenotype estimated with molecular markers and the true genotypic value is stronger than the correlation between the observed phenotype measured directly from the field and the true genotypic value (Wricke and Weber, 1986).

In spite of or maybe because of its importance in breeding, definitions of heritability are numerous and sometimes unclear. Attempts have been made to state a unique definition of heritability. Hanson (1963) defined it as the proportion of genetic gain that is expected to be achieved after selecting for the desired phenotype on a reference unit that can be a single plant, a plot (line), an entry, etc. Broad sense heritability (H) is defined as the fraction of phenotypic variance ascribable to genetic effects, including all kind of genetic sources of variation into a single variance component. Using narrow sense heritability (h^2), breeders attempt to focus only on heritable genetic effects, *i.e.*, those that could be transferred by hybridization, by partitioning the genetic variance into additive variance (that part due to the

effects of individual alleles) and non-additive variance (the part arising from dominance effects within gene pairs, to epistatic interactions between gene pairs, or to higher order epistatic interactions) (Nyquist, 1991). It is important to realize that estimates of variance components and heritability are meaningful only for a particular population, a particular trait, and a particular environment or set of environments tested (Holland *et al.*, 2003). The reference population usually is a random-mating population in Hardy-Weinberg and linkage equilibrium, that is linkage disequilibrium is absent (LD=0). LD is the non-random association of alleles at two or more different loci not necessarily in the same chromosome, numerically LD refers to the difference between observed and expected (with random distribution of alleles) allelic frequencies. In the case of self-pollinated species, such as peanuts, the reference population often is a set of completely inbred genotypes derived from a population in Hardy-Weinberg and linkage equilibrium, which is considered analogous to an F_2 population derived from a cross between two inbred parental lines. The reference population is assumed to have been inbred without selection, random genetic drift, migration or mutation. All of the mentioned factors affect the extent of genetic variability and subsequently the variance and heritability estimates (Falconer and Mackay, 1996).

Molecular Markers

Molecular DNA markers are identifiable specific DNA sequences that may vary within a population, showing different alleles, *i.e.*, polymorphism. They can provide a powerful tool to improve the efficiency of plant breeding methods for quantitative traits (Tanksley,

1983, Varshney *et al.*, 2007). If there are DNA markers associated with regions of the plant genome carrying genes controlling the target trait, then they can be used to select the desired genotypes in a process called marker assisted selection (MAS) (Mohan *et al.*, 1997).

There are essential requirements that must be met for successful application of MAS in plant breeding. First, the markers should co-segregate with or be closely linked (1 cM or less is probably sufficient for MAS) to genes influencing the trait of interest (Mohan *et al.*, 1997). In most cases this co-segregation results from close physical proximity of the gene or genes and the marker. In some cases "cold" recombination spots, regions of the genome where recombination is less likely to occur, are the cause of the phenomenon (Sourdille *et al.*, 2004). A second requirement is that an efficient, economic, reproducible and reasonably easy screening method for the marker should be available. Currently, techniques based on the polymerase chain reaction (PCR) are preferred (Mohan *et al.*, 1997). PCR is a relatively simple and well known technique that requires only small quantities of not very pure DNA. A third requirement is the availability of a method to assess phenotypes accurately and precisely enough so that associations between the trait of interest and the markers can be clearly established. It is important to notice that an extremely accurate and precise method of phenotypic evaluation would possibly signify that there is no need for MAS in the first place. Therefore, a balance between necessity and efficacy of the developed MAS method must be met.

Several screening techniques for molecular markers have been developed and used for the genus *Arachis* including protein based markers or isozymes (Stalker *et al.*, 1994), restriction fragment length polymorphisms (RFLPs; Botstein *et al.*, 1980; Kochert *et al.*,

1991), random amplified polymorphic DNA (RAPDs; Williams *et al.*, 1990; Hilu and Stalker, 1995; Subramanian *et al.*, 2000), and amplified fragment length polymorphism (AFLPs; Vos *et al.*, 1995; Gimenes *et al.*, 2002; Milla *et al.*, 2005). However, little or no variability at the molecular marker level was found among cultivars of *A. hypogaea*, in spite of the existence of pronounced variation in agronomically important traits such as growth habit, maturation time, seed size, etc. As a consequence of the low level of polymorphism found within *A. hypogaea* using the aforementioned marker sytems, the application of molecular markers for breeding purposes, *i.e.*, linkage mapping and posterior MAS, in cultivated peanut is very limited (Hopkins *et al.*, 1999). More recently, simple sequence repeats (SSRs, Powel *et al.*, 1996) have been developed and used in peanuts (Hopkins *et al.*, 1999; He *et al.*, 2003; Varshney *et al.*, 2007; Ferguson *et al.*, 2004; Moretzsohn *et al.*, 2004; Moretzsohn *et al.*, 2005).

Isozymes (Hunter and Merkert, 1957) are proteins with biological activity that perform the same metabolic function but have different sequence of amino acids and/or secondary structure. These differences are evidenced in their electrophoretic mobility. The study of the electrophoretic pattern of these proteins after identifying them for their enzymatic activity or their reaction with specific antibodies constitutes one of the earliest molecular markers (Lacks and Stalker, 1993; Stalker *et al.*, 1994). In spite of being set aside for the more informative modern molecular markers described below, isozymes can still be useful for projects with limited resources and where more powerful differentiation is not necessary.

Restriction fragment length polymorphisms (RFLPs) are identified by the digestion of total DNA by specific restriction enzymes and separation of the produced fragments in an

agarose or polyacrylamide gel by electrophoresis. This process produces a specific pattern in the gel when there are differences in the sequence that is recognized by the cutting enzymes and/or their position in the genome. RFLPs behave as co-dominant markers giving the maximum amount of information obtained from a mapping population. Furthermore, they have a higher degree of reproducibility among laboratories than RAPDs or AFLPs. However, larger amounts of DNA are required for RFLP analysis relative to other systems. Disadvantages associated with using RFLPs as molecular markers include the use of dangerous radioisotopes, large quantities of restriction enzymes, and Southern blotting procedures (Stalker *et al.*, 2001). All of these characteristics made RFLPs not optimal for breeding purposes.

Random amplification of polymorphic DNA (RAPD) is a PCR reaction carried out using primers of random sequence and total genomic DNA. RAPD markers result from DNA sequence variation at primer binding sites and from DNA length differences between primer binding sites (Powell *et al.*, 1996). Using this technique, a random set of primers is used to hybridize with the DNA of the studied organism. Using the same set of primers with different samples, different "patterns" of amplification that can be used for comparison of the studied genomes can be obtained. The technique is appealing because no prior knowledge of the studied genome is required, but it has limitations because large intact sequences of template DNA are required, and it is not highly reproducible among different laboratories.

Sequence characterized amplified regions (SCARs; Paran and Michelmore, 1993) are co-dominant sequence specific markers that are developed from RAPDs. Briefly, the polymorphic RAPD band is isolated from the gel. Afterwards, a PCR reaction, using the

original RAPD primers and the extracted DNA, is carried out. The PCR product is then introduced into a bacterial vector and cloned. Finally, the insert is excised and sequenced in order to create longer primers that are not random any more but are specific for a DNA sequence (Melotto *et al.*, 1996; Naqvi and Chattoo, 1996)

Amplified fragment length polymorphism (AFLP) technique is based on the selective polymerase chain reaction (PCR) amplification. The technique involves digestion of DNA with restriction enzymes and ligation of oligonucleotide adapters. Subsequently, PCR allows selective amplification of sets of restriction fragments. Finally, the PCR products can be separated in an agarose or polyacrylamide gel to detect differences in fragment length (Vos *et al.*, 1995). AFLP markers are reproducible, detect considerable levels of polymorphism, are widely distributed throughout the genome, and do not require previous knowledge of the organism's genome because random sequence primers are used (Lu *et al.*, 1996).

Simple sequence repeat (SSR) markers, also known as microsatellites, are DNA sequences containing tandem repeats of one to six nucleotides (Appleby *et al.*, 2009) that can be present in repeat numbers larger than 30 throughout the cultivated peanut genome. SSRs are highly polymorphic and informative markers, and have been reported to be more variable in *A. hypogaea* than are AFLPs or RAPDs (He *et al.*, 2003). This elevated level of polymorphism originates through mutations that affect the number of repeated units. The screening method for SSRs is based in polymerase chain reaction (PCR), making these markers easy to screen with precision. Moreover, analysis of marker data is amenable to automated allele detection and sizing. SSRs are often multi-allelic and show a high degree of transferability between species as primer sets designed for one species often amplify a

corresponding locus in related species (Appleby *et al.*, 2009). All the mentioned characteristics, especially the hyper-variability among related organisms, make SSRs a suitable marker system for peanut breeding applications including genetic mapping and MAS. Nevertheless, there is an important drawback in the use of SSRs, namely the high cost and difficulty of developing new markers. Use of SSRs requires primers with the exact complementary sequence of the SSR flanking regions, *i.e.*, genome sequence data must be available. Therefore, production of bacterial genomic libraries, DNA sequencing, identification of SSR sequences, and synthesis of primers designed for the neighboring regions of the marker is the preferred method (Hopkins *et al.*, 1999). However, newer techniques with promising results are lowering the cost and effort required to identify new markers. These techniques involve the enrichment of SSR containing sequences prior to the sequencing step (Moretzsohn *et al.*, 2004; He *et al.*, 2003) or the usage of bioinformatics and data mining of existing sequence information to detect SSRs (Moretzsohn *et al.*, 2005). Mainly because these innovative techniques were available, several hundreds of SSR markers have been developed recently, most of them in the last six years (Varshney *et al.*, 2007; Moretzsohn *et al.*, 2004; Moretzsohn *et al.*, 2005; Ferguson *et al.*, 2004; He *et al.*, 2003; Hopkins *et al.*, 1999). However, in spite of having such an array of SSR markers, only one linkage map based on a cultivated-by-cultivated cross has been published to date (Varshney *et al.*, 2009). Therefore, more work needs to be done to evaluate the level of polymorphism in the publicly available SSR markers, and to assess the feasibility of constructing linkage maps and locating QTL in them.

Single nucleotide polymorphisms (SNPs; Shattuck-Eidens *et al.*, 1990) are variations of a single base pair in allele sequences. They are the most common type of sequence difference between alleles (Rafalski, 2002). Platforms and SNP discovery technologies that were created for the human genome project can easily be adopted for crop breeding, making SNPs a very promising alternative for MAS. There are three main ways for SNP discovery: a) re-sequencing of PCR products with or without pre-screening, b) electronic SNP discovery (eSNP), consisting of screening a group of sequences of several different haplotypes and looking for SNPs that occur at a certain frequency, c) eSNP or discovery using databases of expressed sequence tags (EST), that is getting the gene sequence from RNA isolated from living cells and looking for SNPs comparing different haplotypes (Rafalsky, 2002).

Mapping Populations

Mapping populations are necessary for the production of linkage maps because the statistical models used to test linkage require the knowledge of the expected allelic frequencies of the different markers under the assumption of independent assortment (when two loci are in different chromosomes and therefore not linked) in order to identify deviations of the observed data from the expected values and subsequently declare linkage and obtain recombination frequency estimates (Wu *et. al*, 2007). These expected values are known if controlled crosses are performed, first between inbred lines and after that in the subsequent generations. The main types of mapping populations are:

a) F_2 population. Compared to the others, this is the type of mapping population that requires less effort to develop. It is created by selfing the hybrid between two

inbred lines or (in animals or self-incompatible plants) by intermating the progeny of a cross between two more or less inbred parental lines. As stated, it is the easiest population to develop, but has the disadvantage of being an ephemeral population. An F_2 population has limited plant material from which to isolate DNA or take phenotypic measurements from because once this generation is used there is no way to generate an exact replicate of it unless the species is amenable to cloning. As a consequence, this population cannot be evaluated in replicated trials over locations and years, making it difficult to obtain accurate and precise estimates of phenotypic values. Furthermore, linkage maps based on these populations are based on a single recombination event for a particular locus during production of F_1 gametes; therefore, this population has limited recombination information (Wu *et. al*, 2007).

b) F_2-derived F_3 ($F_{2:3}$). Is obtained by selfing individuals of an F_2 population for one generation. This population allows one to determine the genotype of the F_2 by pooling the DNA of plants of the F_3 progeny.

c) Backcrosses. Backcross populations (BC_1F_1) are formed by crossing an F_1 back to either of its parents. In self-pollinated species, this type of population requires more effort to develop than does an F_2 because more artificial hybridizations are required. Its main advantage is the possibility of using the population directly for backcross breeding. However, in populations of equal sizes an F_2 has twice as much linkage information as an BC_1F_1, because in the case of the F_2 both gametes (male and female) are sampled, whereas, for BC_1F_1 only the non-recurrent parent

gamete (F_1) is sampled. Furthermore, as is the case with F_2 populations, backcross populations are ephemeral and cannot be exactly replicated unless the species can be cloned (Wu *et. al*, 2007).

d) Double haploids (DHs). The development of this population typically involves the employment of highly technical protocols, such as anther culture. As an F_2 they only carry the information of a single recombination event. They allow for instant creation of homozygous lines, saving a lot of time when compared to other mapping populations. They are permanent populations that can be replicated over years and environments (Wu *et. al*, 2007).

e) Near-isogenic lines (NILs). They are generated by repeated backcrossing of the F_1 with a recurrent parent. They consist of lines that are genetically identical except for the genes of interest and are permanent mapping populations that allow for replication. However, they require many generations for development (Wu *et. al*, 2007).

f) Recombinant inbred lines (RILs). These can be derived either by repeated selfing or sibling mating of the progeny of an F_1 from a cross between two inbred lines. RILs usually show a high degree of homozygosity at all loci because they originate from continuous inbreeding for several generations. Assuming that the rates of spontaneous mutation and outcrossing are negligible, RILs can be maintained by self-fertilization, allowing the replication of identical genotypes at the scale of time and space that is required for a particular genetic and/or biological study. Therefore, RILs can serve as a permanent mapping population

for multiple uses because they are fixed and homozygous for two alternative alleles at all genes. Some lines are the same as parental (non-recombinant) types, whereas others are recombinant types (Wu *et. al.*, 2007). RILs accumulate crossovers that occur at each meiosis with every generation, for the loci that remain in heterozygous state, and thus the proportion of recombinant zygotes in RILs, *i.e.* the probability that two linked loci have different parental alleles, is higher than it would be in an F_2 (Wu *et. al.*, 2007). Moreover, this additional opportunity for crossovers tends to produce larger estimated genome sizes and, as a consequence, RILs could require larger population sizes to equal the efficiency for detecting QTLs of other population development methods. However, the ability to replicate the RILs over different environments (years and/or locations) and the resulting increase in the precision of phenotypic value estimates normally compensate for the larger estimated genome size. As a result, RILs typically have been more efficient in detecting QTLs than equal numbers of backcross progeny (Lander and Botstein, 1989). However, the greater amount of time and effort required to develop RILs have to be taken into consideration when choosing a mapping population.

Linkage map construction

Linkage is defined as the tendency for genes or loci located in close physical proximity on the same chromosome to be passed together into the next generation (Haldane,

1931). Linkage mapping is based on the analysis of this co-segregation in genetic markers, mostly molecular DNA markers in recent years (Wu *et al.,* 2007). Highly penetrant genes controlling easily classified phenotypes can also be used as genetic markers, however, the overwhelming majority of modern linkage mapping studies use molecular DNA markers. In order to be able to distinguish between recombinant and non-recombinant progeny, one must know the allelic composition of the parental lines. That is, the *linkage phases* of the markers used for mapping must be known. Subsequently, the strength of the linkages among the markers can be expressed in terms of recombination frequencies or genetic distances corrected using mapping functions based on assumptions regarding chromosomal behavior (Wu *et al.,* 2002). Finally, the relative positions of two or more markers can be defined and a linkage map created. The previous was a simplified version of how to conduct linkage mapping; in the following paragraphs we will discuss some of the details that this process entails.

The first step in mapping is to analyze the segregation pattern of the individual markers used to construct the map. One of the assumptions of linkage mapping is that individual markers follow Mendelian segregation ratios because the populations that were genotyped are supposed to have been developed in the absence of any force that could disturb this ratio, *e.g.*, selection, migration, mutation or genetic drift. This is a basic assumption because only after the nature of the single marker ratios is known can appropriate statistical methods be identified and a meaningful linkage map be produced. Generally, the methods of choice to evaluate segregation patterns are the Pearson chi-square test or the likelihood ratio chi-square (Wu *et. al.*, 2007).

The second step in mapping includes the analysis of linkage among markers and the estimation of recombination fractions. For this purpose, there are two different statistical approaches. In two-point analysis, only two markers are evaluated at a time and the only parameter estimated is the recombination frequency (*r*) between these two markers. In three-point analysis three recombination frequencies must be considered because there are three pair-wise combinations of three markers evaluated. Three-point analysis may provide increased precision of the recombination frequencies when the data include markers that are not informative. Furthermore, it allows one to determine the order of markers (Thompson, 1984; Wu *et. al.*, 2002).

Map distance is defined as the expected number of crossovers occurring between two loci on a single chromatid during meiosis (Wu *et al.*, 2007). Thus, the following step involves the conversion of recombination frequencies (*r*) into genetic map distances. Recombination frequencies in a series of regions are not additive because there can be numbers of crossover events that would yield the same observed *r* as long as they occur in an even number. Genetic map distances are additive. To accomplish the conversion one must use a map function, a mathematical function that uses certain assumptions to estimate the map distance from the number of observed recombinants. Map functions have several variants including the following:

(1) Mather's function assumes that the recombination fraction (*r*) between two loci is half the probability of chiasmata occurring in all four strands of tetrads between the loci (Mather, 1938),

(2) Morgan's function assumes that there is at most a single crossover happening on the interval of two loci and that a crossover on an interval is proportional to the map length of the interval (Morgan, 1928),

(3) Haldane's function assumes that crossover events occur randomly and are independent from each other (Haldane, 1919). Experimental results suggest that cross over events are not independent especially when we consider loci that are located in close physical proximity to each other,

(4) Kosambi's function (Kosambi, 1994) is derived from a generalization of the Haldane function including the notion of interference. Interference is defined as the tendency of not finding two crossing over events in close proximity to each other. In other words, the formation of a chiasma at meiosis tends to inhibit further crossing over in its vicinity (Wu *et al*., 2007). The value of interference measures the deviation of observed recombinations in different intervals from the expected recombinations that are supposed to happen if the processes were independent between intervals (Muller, 1916). In practice, interference appears to be close to absolute when the evaluated intervals are very close to each other, *i.e.*, recombination fractions seem to be purely additive at short distances, and seem to be non-existent when the considered intervals are far away from each other in which case Kosambi's function is identical to Haldane's.

There are only two very recent research projects, Varshney *et al.* (2009) and Hong *et al.* (2010), that produced linkage maps for populations developed using cultivated-by-cultivated peanut crosses (*A. hypogaea* x *A. hypogaea*). Both of them used SSR markers for their major

advantages, *i.e.*, a high degree of reproducibility, co-dominance, transferability, and the possibility for comparison between maps when the same set of markers are used (Hong *et al.*, 2010). The first linkage map for a cultivated-by-cultivated cross was constructed by Varshney *et al.* (2009) using a set of 135 informative polymorphic SSR markers and 318 RILs in the $F_{8:9}$ generation. Mapmaker Macintosh version 2.0 (Lander *et al.*, 1987) and Kosambi's map function were used for map construction. The markers were included in two rounds. First, they used markers that did not show segregation distortion using a highly stringent statistical test to produce a highly certain "backbone" or scaffold of markers, and later filled in the map with other markers with a less stringent test. A linkage map with 22 linkage groups (LG) and a total map distance of 1270.5 cM was produced. Varshney *et al.* (2009) mapped some quantitative traits (transpiration, transpiration efficiency, specific leaf area) using the composite interval mapping (CIM) method in QTL Cartographer, v 2.5 (Wang *et al.*, 2007). They identified several QTL for each trait. However, the QTL explained very little of the total phenotypic variation and were not deemed useful in application of MAS breeding methods. Possible explanations for that result are the low marker density obtained and the lack of phenotypic variability in the RILs for the evaluated traits.

The latest attempt to produce a cultivated peanut linkage map that could allow for molecular breeding applications was published by Hong *et al.* (2010). In this study, three different RIL populations sharing the female parent, a spanish-type genotype (cultivar Yueyou 13), and a total of 192 SSR markers that were polymorphic for at least one of the populations were used. The populations had sizes of 142, 136, and 84 lines (146, 64, and 124

markers were polymorphic for each of the mentioned populations). The non-shared parental genotype (used as male) were a virginia type with high protein content, cultivar Zhenzhuhei, for the population with 142 lines; a spanish type with reported resistance to aflatoxin contamination, cultivar J11, for the population with 136 lines; and a spanish type with high oil content, cultivar Fu 95-5, for the population with 84 lines.

JoinMap 3.0 (Van Oojen and Voorips, 2001) was used for linkage map construction. A linkage map was created for each of the populations, with 19, 21, and 13 linkage groups respectively. The individual maps covered 684.9, 540.69, and 401.7 cM of the peanut genome. Segregation distortion was detected in 22.8%, 13.6%, and 8.5% of the polymorphic markers for the three populations, respectively. Possible explanations for the high frequency of segregation distortion are chromosome loss, genetic isolation, the presence of viability genes, or the inadvertent application of selection for one parental allele (Hong *et al.*, 2010). Subsequently, a composite genetic linkage map was constructed using the individual maps. Briefly, linkage groups containing common markers in individual maps were assigned onto a single integrated linkage group. The composite map used 175 marker loci, produced 22 linkage groups and covered 885.4 cM of the peanut genome. The mean interval between adjacent markers was 5.79 cM and 85% of the intervals were smaller than 10cM. The map distance obtained by Hong *et al.* is considerably smaller than the one obtained by Varshney *et al.* (885.4 cM vs 1270.5 cM), or the two available maps for A genome diploid *Arachis* species (1063 cM and 1230.89 cM) (Moretzsohn *et al.*, 2005; Halward *et al.*, 1993) and even smaller when compared to the map of a synthetic tetraploid with RFLP markers (2210 cM) (Burow *et al.*, 2001). A deficient saturation of the SSR markers used in the cultivated peanut

studies could contribute to that observation (370 RFLP markers used in the synthetic polyploid vs. 175 and 135 SSR markers in the cultivated peanut studies). Another important difference is the computer program used to construct the linkage maps in each study. Linkage maps constructed with Mapmaker tend to be consistently larger than the ones constructed with JoinMap even when both programs use the Kosambi mapping function (Kosambi, 1994). Mapmaker uses the multilocus-likelihood method to create linkage maps, and this method assumes no crossover interference. JoinMap's method of map construction includes a parameter for interference, and as a result the maps created by JoinMap tend to be smaller, which is a better approximation of natural conditions when interference is present as is the most likely the case in peanuts (Hong *et al.*, 2010).

Attempts have been made to correlate the present cultivated peanut maps with previous maps obtained from crosses involving synthetic allotetraploids (Moretzsohn *et al.*, 2005), diploid *Arachis* genomes, or even other members of the leguminosae family (Varshney *et al.*, 2009). Synteny of small regions of linkage groups among the previous maps and the newer ones have been found (Varshney *et al.* 2009; Hong *et al.*, 2010). However, more conclusive results and better comparisons will require the effort of similar projects using comparable markers (SSR), and the use of newer marker technologies (SNPs) that will allow for greater marker density.

Quantitative Trait Loci (QTL)

Until recently, it was very difficult to study individual gene inheritance patterns and other important properties such as gene frequencies and degree of the effects of individual

genes for a quantitative trait of interest. As a consequence, many quantitative genetic theories required unrealistic assumptions, or overgeneralizations, about the genes influencing quantitative traits. These assumptions included: that the gene frequencies of the genes involved are the same (true in the case of F_2 population), that the genes' effects and dominance relations are very close to be equal across all genes, or even that there is an infinite number of loci governing the trait of interest (Falconer and Mackay, 1996). However, new methods such as molecular markers are now available for the study of the individual loci and sometimes individual genes influencing quantitative traits. These loci, called "quantitative trait loci" or "QTL", may represent a segment of a chromosome including multiple genes. Even in that case, QTL studies allow for a more realistic approximation for better models in quantitative genetics, could allow for the employment of transgenic technologies in quantitative traits, and may improve the efficiency of traditional breeding methods through MAS. Methods of detection of QTL are based on the presence of linkage disequilibrium ($LD \neq 0$) between molecular marker loci and the QTL. In other words, there is a tendency for co-segregation of a certain allele for the marker loci (identifiable through genotyping) and certain allele of the QTL, producing a particular measurable phenotype. Moreover, to successfully detect QTL one needs a linkage map with adequate marker density and good coverage of the entire genome of the organism under study. If these conditions are met, then one is likely to find some markers with strong associations with the QTL. Ideally molecular markers for QTL studies should be highly polymorphic to allow construction of a good linkage map, abundant enough to get good marker density and genome coverage, co-dominant so as to identify all possible genotypes, and not influenced by

selection or other evolutionary forces that could affect the observed allelic frequencies (Falconer and Mackay, 1996). Another important requirement is that the population under study shows sufficient variability of the trait controlled by the QTL. Furthermore, at the time of choosing the parents for the mapping population it is better to choose genotypes that have fixed alleles for both the marker loci and the QTL. These alleles ought to differ between parental genotypes. It is also advisable that the alleles influencing the QTL are in "association", that is that the loci increasing (or decreasing) the phenotypic value of the studied trait should be together in the same parental genotype. QTL detection is less probable and mapping is made more difficult when there is a mixture of alleles that both increase and decrease the final value of the trait, *i.e.*, when the alleles are "dispersed" across the two parents (Falconer and Mackay, 1996).

The most commonly used mapping populations are F_2 and backcross populations. The backcross design is considered less powerful than the F_2 design because backcrosses using one parent detect only effects of heterozygous genotypes. These effects are just one half of the homozygous effects that can be detected using an F_2 (Falconer and Mackay, 1996). Backcrossing using both parents allows for estimation of the effects of homozygous loci but is less efficient than using an F_2. Another way to see the disadvantage of backcrossing (compared to F_2 design), is that in each backcrossed individual only a single gamete from the F_1 is represented, whereas, in the F_2 design both gametes are involved. Using an F_2 population, we can sample twice as many potentially recombinant gametes as in a comparable backcrossed population, thus increasing the statistical power of the design.

In self-pollinated species in which population development can include several (in theory infinite) generations of inbreeding, there is the option of using recombinant inbred lines (RILs). Another alternative is using randomly mated populations. With this method, it is only possible to detect QTL with extremely strong linkage with the markers used. Otherwise, mildly or weakly linked loci will be likely to be found in linkage equilibrium (LD=0).

The basic principle used for QTL detection is to find differences in the mean phenotypic value for the trait of interest among the different marker alleles (or classes). This principle can be applied to individual markers (as in single marker analysis), or group of markers (as in interval mapping analysis or composite interval mapping analysis). The resolution of most mapping projects is about 20cM. Therefore, in many cases the number of QTL is underestimated because two or more loci can be included as a single QTL (Falconer and Mackay, 1996). However, fine mapping techniques based in an increased number of meiotic events (*i.e.* increased population sizes) and higher marker density (*i.e.* more markers and/or more equally distributed), can decrease the resolution to about 3cM allowing for a more close approximation to the real number of QTL.

The most basic method for analyzing data and detecting QTL is single marker analysis. According to this method, a statistically significant difference in the phenotypic value for the trait between marker genotypic classes (*i.e.* AA, aa, or Aa) is taken as evidence of linkage between the QTL and the marker locus. In spite of its evident simplicity, there is a major drawback in using single marker analysis: QTL effects are affected by recombination frequencies. For example, we would not be able to distinguish between a QTL with a modest

effect that is very close to the marker, and a QTL with a large effect that is farther away from the marker locus. Interval mapping analysis helps to eliminate the mentioned confounding effect by simultaneously considering pairs of linked marker loci. Estimates of the QTL effect are given for a particular map position relative to the flanking markers. Compared with single marker analysis and assuming that the recombination frequency between flanking markers is negligible, interval mapping allows for use of much smaller population sizes to detect QTL of the same magnitude (Lander and Botstein, 1989).

The methods of choice for detecting QTL are based on the maximum likelihood function because they provide a better approximation to the real properties and distributions of phenotypes across marker classes than do other possible detection methods that assume normal distribution, *e.g.*, t-tests or analysis of variance (Lander and Botstein, 1989). Briefly, the maximum likelihood procedure uses the observed data (number of individuals and their phenotype for each marker class) and a set of unknown parameters (recombination frequency, variances of QTL genotypes) to define a function. Then, an iterative computer program (*e.g.* QTL-Cartographer, QTL-Mapmaker) tries different random values for the unknown parameters and calculates the likelihood function (L) for each of the different sets of random values. The tested parameter values that maximize the likelihood function are selected as the parameter estimates. In other words, the parameter values that maximize the probability of getting the observed data are selected (Lander and Botstein, 1989). The test statistic used is the $\log_{10}$ of the ratio L/L_0 which is distributed as a chi-square statistic where L is the obtained likelihood and L_0 is the estimated likelihood function when no QTL is present (Falconer and Mackay, 1996). This statistic reflects the probability of obtaining the observed

data when a QTL is present compared with assuming it is absent (Lander and Botstein, 1989). When interval mapping is used, the mentioned $\log_{10}$ likelihood ratio is called the LOD or "logarithm of odds" score. In this case, LOD scores are estimated for different positions between marker intervals and the parameter estimates are selected when the LOD score is maximized. The threshold LOD value above which a QTL is detected is determined by the Type I error rate (usually α=0.05) and the number of marker intervals tested. Both factors are ultimately determined by the genome size, number of markers used (marker density), and the number of chromosomes of the studied organism. However, LOD scores usually range from 2 to 3 which in most cases corresponds to a Type I error rate of around 0.0001 for each interval tested (Lander and Botstein, 1989).

QTL studies in *A. hypogaea* have been hindered because of the lack of linkage maps with enough marker density and adequate resolution. However, the recent increased availability of highly variable markers (such as SSR) makes the construction of linkage maps and QTL studies possible in cultivated peanuts. Khedikar *et al.* (2010) used 268 RILs and 67 polymorphic SSR markers to produce a partial linkage map with 14 linkage groups. Phenotypic data for late leaf spot (LLS) and rust, two major leaf diseases, were collected for three years. Afterwards, composite interval analysis (CIM; Zeng, 1994) identified 11 QTL associated with LLS explaining 1.7-6.5% of the phenotypic variation and 12 QTL associated with rust explaining 1.7-55.2% of the phenotypic variation. One marker that contributed 6.9-55.2% of the phenotypic variation for rust was identified using a wide range of resistant and susceptible genotypes. This marker, QTL(rust)01, can be used for introgression purposes using marker-assisted backcrossing.

Varshney *et al.* (2009) used 318 RILs derived from crosses between cultivated peanut lines and a total of 144 polymorphic markers to produce a genetic map with 22 linkage groups. Phenotypic data was obtained for drought-related traits including transpiration (T), transpiration efficiency (TE), specific leaf area (SLA), and SPAD chlorophyll meter reading (SCMR). All phenotypic data was collected for one location for two consecutive years, 2004 and 2005. Subsequently, CIM analysis (Zeng, 1993; 1994) using the phenotypic and genotypic data yielded 4 QTL for T, 4 for TE, 14 for SLA, and 6 for SCMR. However, the phenotypic variation explained by these QTL was low, ranging from 3.5-14.1%. Therefore, none of the identified QTL has potential for use in MAS. There are two likely explanations for the low phenotypic variation explained by each QTL: low phenotypic variation was observed among the RILs measured in this study, and insufficient marker density hindered detection of QTL with larger phenotypic effects.

Peanut maturity is an economically important quantitative trait with complex heredity patterns for which an alternative selection method, other than direct phenotypic selection, would be highly advantageous. In recent years, there has been an increased interest in developing molecular markers, mainly SSRs, and genomic data for cultivated peanuts. Thanks to the increasing availability of this data and molecular biology tools, MAS for early maturity is an attractive option as an approach to improve breeding methods for early maturity.

REFERENCES

Appleby, N., D. Edwards, and J. Batley. 2009. New technologies for ultra-high throughput genotyping in plants. Chapter 2, p. 19-39. *In* D.J. Somers, P. Langridge, and J.P. Gustafson (eds.), Methods in Molecular Biology, Vol 513: Plant genomics methods and protocols. Humana Press, New York, NY.

Bell, M. J., R.C. Roy, M. Tollenaar, and T.E. Michaels. 1994. Importance of variation in chilling tolerance for peanut genotypic adaptation to cool, short-season environments. Crop Sci. 34: 1030-1039.

Botstein, D., R.L. White, M. Skolnick, and R.W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32: 314-331.

Burow, M.D., C.E. Simpson, J.L. Starr, and A.H. Patterson. 2001. Transmission genetics of chromatin from a synthetic amphidiploids to cultivated peanut (*Arachis hypogaea* L.): broadening the gene pool of a monophyletic polyploidy species. Genetics. 159: 823-837.

Davidson, J.I., T.B. Whitaker, and J.W. Dickens. 1982. Grading, cleaning, storage, shelling, and marketing of peanuts in the United States, p. 571-623. *In* H.E. Pattee and C.T. Young (eds.), Peanut Science and Technology. Am. Peanut Res. Educ. Soc., Inc., Yoakum, TX.

Dudley, J.W. 1993. Molecular markers in plant improvement: manipulation of genes affecting quantitative traits. Crop Sci. 33: 660-668.

Falconer, D.S., and T.F. Mackay. 1996. Continuous variation: metric characters, p. 100-104. *In* Longman A.W. (Ed.), Quantitative Genetics. Pearson Education Limited, Essex, UK.

Ferguson, M.E., M.D. Burow, *S*.R. Schulze, P.J. Bramel, A.H. Paterson, S. Kresovich, and S. Mitchell. 2004. Microsatellite identification and characterization in peanut (*A. hypogaea* L.). Theor. Appl. Genet. 108: 1064–1070.

Gimenes, M.A., C.R. Lopes, and J.F.M. Valls. 2002. Genetic relationships among *Arachis* species based on AFLP. Genet. Mol. Biol. 25: 349-353.

Grimm, C.C., E.T. Champagne, and T.H. Sanders. 1998. Determination of peanut maturity using a hunter colorimeter. Peanut Sci. 25: 99-103.

Hammons, R. O. 1982. Origin and early history of the peanut, p. 1-20. *In* H.E. Pattee and C.T. Young (eds.), Peanut Science and Technology. Am. Peanut Res. Educ. Soc., Inc., Yoakum, TX.

Haldane, J.B.S. 1919. The combination of linkage values and the calculation of distances between the loci of linked factors. J Genet. 8: 299–309.

Haldane, J.B.S., and C.H. Waddington. 1931. Inbreeding and linkage. Genetics 16: 357-374.

Halward, T. M., H.T. Stalker, and G. Kochert. 1993. Development of an RFLP linkage map in diploid peanut species. Theor. Appl. Genet. 87: 379-384.

Hanson, W.D. 1963. Heritability, p. 125-140. *In* W.D. Hanson and H.F. Robinson (eds). Statistical Genetics and Plant Breeding, Publication 982, National Academy of Sciences-National Research Council, Washington D.C.

He, G., R. Meng, M. Newman, G. Gao, R.N. Pittman, and C.S. Prakash. 2003. Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L). BMC Plant Biol. 3: 3.

http://www.biomedcentral.com/1471-2229/3/3.

Hilu, K.W. and H.T. Stalker. 1995. Genetic relationships between peanut and wild species of *Arachis* sect. *Arachis* (*Fabaceae*): evidence from RAPDs. Plant. Syst. Evol. 198: 167-178.

Holland, J.B., W.E. Nyquist, and C.T. Cervantes-Martínez. 2003. Estimating and interpreting heritability for plant breeding: an update. Plant Breed. Rev. 22: 9–111.

Holland, J.B. 2007. Genetic architecture of complex traits in plants. Curr. Opin. Plant Biol. 10: 156-61.

Hong, Y., X. Chen, X. Liang, H. Liu, G. Zhou, S. Li, S. Wen, C.C. Holbrook, and B. Guo. 2010. A SSR-based composite genetic linkage map for the cultivated peanut (*Arachis hypogaea* L.) genome. BMC Plant Biol. 10: 17. http://www.biomedcentral.com/1471-2229/10/17

Hopkins, M.S., A.M. Casa, T. Wang, S.E. Mitchell, R.E. Dean, G.D. Kochert, and S. Kresovich. 1999. Discovery and characterization of polymorphic simple sequence repeats (SSRs) in peanut. Crop Sci. 39: 1243-1247.

Hunter, R. L., and C.L. Merkert. 1957. Histochemical demonstration of enzymes separated by zone electrophoresis in starch gels. Science 125: 1294-1295

Isleib, T. G., and J. C. Wynne. 1992. Use of the plant introductions in peanut improvement, pp. 75-116. *In* H. L. Shands and L. E. Weisner (eds.) Use of Plant Introductions in Cultivar Development, Part 2. CSSA Special Publ., Madison. WI.

Isleib, T.G., H.E. Pattee, and P.W. Rice. 1997. A laboratory colorimeter method to measure pod brightness in virginia-type peanuts. Peanut Sci. 24: 81-84.

Isleib, T. G., C. C. Holbrook, and D. W. Gorbet. 2001. Use of peanut introductions in peanut cultivar development. Peanut Sci. 28:96-113.

Knauft, D. A., A. J. Norden, and D. W. Gorbet. 1987. Peanut breeding, p. 346-384. *In* W.R., Fehr (ed.), Principles of Cultivar Development. Vol. 2. Macmillan Publ. Co., New York, NY.

Khedikar, Y.P., M.V.C. Gowda, C. Sarvamangala, K.V. Patgar, H.D. Upadhyaya, and R.K. Varshney. 2010. A QTL study on late leaf spot and rust revealed one major QTL for molecular breeding for rust resistance in groundnut (*Arachis hypogaea* L.). Theor. Appl. Genet. 121: 971-984.

Kochert, G., T. Halward, W.D. Branch, and C.E. Simpson. 1991. RFLP variability in *p*eanut (*Arachis hypogaea*) cultivars and wild species. Theor. Appl. Genet. 81: 565-570.

Kosambi, D.D. 1994. The estimation of map distance from recombination values. Ann Eugen. 12:172-175.

Krapovickas, A. and W.C. Gregory. 1994. Taxonomy of the genus *Arachis* (Leguminosae). Bonplandia 8: 1-186.

Lacks, G.D., and H.T. Stalker. 1993. Isozyme analyses of *Arachis* species and interspecific hybrids. Peanut Sci. 20: 76-81

Lander, E.S., and D. Botstein. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics. 121: 185-199.

Lander, E.S., P. Green, J. Abrahamson, A. Barlow, and M. J. Daly. 1987. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174-181.

Lu, J., M.R. Knox, M.J. Amrose, J.K.M. Brown, and T.H.N. Ellis. 1996. Comparative analysis of genetic diversity in pea assessed by RFLP- and PCR-based methods. Theor. Appl. Genet. 93: 1103-1111.

Mather, K. 1938. Crossing-over. Biol Reviews Camb Phil Soc 13:252–292.

Melotto, M., L. Afanador, and J.D. Kelly. 1996. Development of a SCAR marker linked to the I gene in common bean. Genome 39: 1216-1219.

Milla, S.R., T.G. Isleib, and H.T. Stalker. 2005. Taxonomic relationships among *Arachis* sect. *Arachis* species as revealed by AFLP markers. Genome 48: 1-11.

Mohan, M., S. Nair, A. Bhagwat, T.G. Krishna, M. Yano, C.R. Bhatia, and T. Sasaki. 1997. Genome mapping, molecular markers and marker-assisted selection in crop plants. Molecular Breeding 3: 87-103.

Moretzsohn, M.C., M.S. Hopkins, S.E. Mitchell, S. Kresovich, J.F.M. Valls, and M.E. Ferreira. 2004. Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. BMC Plant Biol. 4: 11. http://www.biomedcentral.com/1471-2229/4/11.

Moretzsohn, M.C., L. Leoi, K. Proite, P.M. Guimaraes, S.C.M. Leal-Bertioli, M.A. Gimenes, W.S. Martins, J.F.M. Valls, D. Grattapaglia, and D.J. Bertioli. 2005. A microsatellite-based, gene-rich linkage map for the AA genome of *Arachis* (*Fabaceae*). Theor. Appl. Genet. 111: 1060-1071.

Morgan, T.H. 1928. The Theory of Genes. Yale University Press, New Haven, CT.

Moss, J.P., and V.R. Rao. 1995. The peanut-reproductive development to plant maturity, p 1-13. *In* H.E. Pattee and H.T. Stalker (eds.), Advances in Peanut Science. Am. Peanut Res. Educ. Soc., Inc., Stillwater, OK.

Muller, H.J. 1916. The mechanism of crossing-over. Am. Nat. 50: 193-221.

Naqvi, N.I., and B.B. Chattoo. 1996. Development of a sequence characterized amplified region (SCAR) based indirect selection method for a dominant blast-resistance gene in rice. Genome 39: 26–30. doi:10.1139/g96-004.

Nigam, S.N., and R. Aruna. 2008. Improving breeding efficiency for early maturity in peanut, p. 295-322 *In*: J. Janick (ed.). Plant Breeding Rev., *vol 30.* John Wiley & Sons, Inc., Hoboken, NJ.

Nyquist, W.E. 1991. Estimation of heritability and prediction of selection response in plant populations. Crit. Rev. Plant Sci. 10: 235-322.

Paran, I., and R.W. Michelmore. 1993. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. Theor. Appl. Genet. 85: 985-993.

Pattee, H.E., E.B. Johns, J.A. Singleton, and T.H. Sanders. 1974. Composition changes of peanut fruit parts during maturation. Peanut Sci. 1: 57–62.

Powell, W., M. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey, and A. Rafalski. 1996. The comparison of RFLP, RAPD, AFLP, and SSR (microsatellite) markers for germplasm analysis. Mol. Breeding 2: 225-238.

Rafalski, A. 2002. Applications of single nucleotide polymorphisms in crop genetics. Current Opinion in Plant Biol. 5: 94–100.

Rowland, D.L., R.B. Sorensen, C.L. Butts, and W.H. Faircloth. 2006. Determination of maturity and degree day indices and their success in predicting peanut maturity. Peanut Sci. 33: 125–136.

Sanders, T.H., A.M. Schubert, and H.E. Pattee. 1982a. Maturity methodology and postharvest physiology, p. 624-654. *In* H.E. Pattee and C.T. Young (eds.) Peanut Science and Technology. Am. Peanut Res. Educ. Soc., Inc., Yoakum, Texas, USA.

Sanders, T.H., J.A. Lansden, R.L. Greene, J.S. Drexler, and E.J. Williams. 1982b. Oil characteristics of peanut fruit separated by a non-destructive maturity classification method. Peanut Sci. 9: 20-23.

Sanders, T. H. 1989. Maturity distribution in commercially sized Florunner peanuts. Peanut Sci. 16: 91-95.

Sanders, T.H., J.R. Vercellotti, P.D. Blankenship, K.L. Crippen, and G.V. Civille. 1989. Effect of maturity on roast color and descriptive flavor of peanuts. J. Food Sci. 54: 475-477.

Shattuck-Eidens, D.M., R.N. Bell, S.L. Neuhausen, and T. Helentjaris. 1990. DNA sequence variation within maize and melon: observations from polymerase chain reaction amplification and direct sequencing. Genetics 126: 207-217.

Singh, A.K. 1985. Genetic introgression from compatible wild species into cultivated groundnut. *In* Proceedings of the International Workshop on Cytogenetics of *Arachis*. ICRISAT Press, Patancheru, A.P., India. p. 107.

Singh, A.K., and R.W. Gibbons. 1986. Wild species in crop improvement: Groundnut – a case study. *In* P.K. Gupta and J.R. Bahl (eds.) Advances in genetics and crop improvement. Rastogi and Co., Meerut, India.

Sing, B., and U. Singh. 1991. Peanut as a source of protein for human foods. Plant Foods for Human Nutrition 41: 165-177.

Sourdille, P., S. Singh, T. Cadalen, G.L. Brown-Guedira, G. Gay, L. Qi, B.S. Gill, P. Dufour, A. Murigneux, and M. Bernard. 2004. Microsatellite-based deletion bin system for the establishment of genetic-physical map relationships in wheat (*Triticum aestivum* L.). Funct. Integr. Genomics 4: 12–25.

Stalker, H.T., T.D. Phillips, J.P. Murphy, and T.M. Jones. 1994. Variation of isozyme patterns among *Arachis* species. Theor. Appl. Genet. 87: 746-755.

Stalker, H. T., M. K. Beute, B. B. Shew, T. G. Isleib and G. Kochert. 1995. Unlocking the peanut genome, pp. 319-334. *In* J. H. Williams, D. G. Cummins, G. Hutto and A. King (eds.) Impacts and Scientific Advances through Collaborative Research on Peanut. Proc. Peanut CRSP workshop. March 29 - 31, 1995. Arlington, VA.

Stalker, H.T., T. Halward, and G. Kochert. 2001. RFLP map of peanut, p. 285-299. *In* R. L. Phillips and I.K. Vasil (eds.). DNA-Based Markers in Plants. Kluwer Academic Publishers. Printed in the Netherlands.

Subramanian, V., S. Gurtu, R.C. Nageswara Rao, and S.N. Nigam. 2000. Identification of DNA polymorphism in cultivated groundnut using random amplified polymorphic DNA (RAPD) assay. Genome, 43: 656-660.

Tanksley, S.D. 1983. Molecular markers in plant breeding. Plant Mol. Biol. Rep. 1:3-8.

Tapiero, H., G. Nguyen, P. Couvreur , and K.D. Tew. 2002. Polyunsaturated fatty acids (PUFA) and eicosanoids in human health and pathologies. Biomed Pharmacother. 56: 215–222.

Thompson, E.A. 1984. Information gain in joint linkage analysis. IMA Journal of Mathematics Applied in Medicine and Biology 1: 31-49.

Upadhyaya, H.D., L.J. Reddy, C.L.L. Gowda, and S. Singh. 2006. Identification of diverse groundnut germplasm: Sources of early maturity in a core collection. Field Crops Res. 97: 261–271.

USDA-FAS Foreign Agricultural Service. 2011a. Table 01: Major Oilseeds: World Supply and Distribution. February, 2011a. Available at http://www.fas.usda.gov/psdonline/psdreport.aspx?hidReportRetrievalName=BVS&hidReportRetrievalID=700&hidReportRetrievalTemplateID=5

USDA-FAS Foreign Agricultural Service. 2011b. Table 13: Peanut Area, Yield, and Production. Available at http://www.fas.usda.gov/psdonline/psdReport.aspx?hidReportRetrievalName=Table+13+Peanut+Area%2c+Yield%2c+and+Production&hidReportRetrievalID=918&hidReportRetrievalTemplateID=1

USDA-NASS National Agricultural Statistics Service. 2011a. Peanut Stocks and Processing. Available at http://usda.mannlib.cornell.edu/usda/current/PeanStocPr/PeanStocPr-02-25-2011.pdf

USDA-NASS National Agricultural Statistics Service. 2011b. Crop Values 2010 Summary.

Available at http://usda.mannlib.cornell.edu/usda/current/CropValuSu/CropValuSu-02-16-2011.pdf

Van Oojen, ,J.W., and R.E. Voorips. 2001. *JoinMap*® Version 3.0, Software for the calculation of genetic linkage maps. Plant Research International, Wageningen. The Netherlands.

Varshney, R.K., D.A. Hoisington, H.D. Upadhyaya, P.M. Gaur, S.N. Nigam, K. Saxena, V. Vadez, N. K. Sethy, S. Bhatia, R. Aruna, M.V. Channabyre Gowda, and N.K. Singh. 2007. Molecular genetics and breeding of grain legume crops for the semi arid tropics, p. 207-242. *In*: R.K. Varshney, and R. Tuberosa (eds) Genomics-Assisted Crop Improvement, *vol 2* : Genomics Applications in Crops , Springer, Dordrecht, The Netherlands.

Varshney, R.K., D.J. Bertioli, M.C. Moretzsohn, V. Vadez, L. Krishnamurthy, R. Aruna, S.N. Nigam, B.J. Moss, K. Seetha, K. Ravi, G. He, S.J. Knapp, and D.A. Hoisington. 2009. The first SSR-based genetic linkage map for cultivated groundnut (*Arachis hypogaea* L.). Theor. Appl. Genet. 118: 729-739.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res. 23: 4407-4414.

Wang, S., C.J. Basten, and Z.B. Zeng. 2007. Windows QTL Cartographer 2.5. http://statgen.ncsu.edu/qtlcart/WQTLCart.htm.

Williams, E.J., and J.S. Drexler. 1981. A non-destructive method for determining peanut pod maturity. Peanut Sci. 8: 134-141.

Williams, J.G.K., A.R. Kubelik, K.J. Livak, J.A. Rafalski, and S.V. Tingey. 1990. DNA polymorphism amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18: 6531-6535.

Wu, R.L., C.X. Ma, I. Painter, and Z.B. Zeng. 2002. Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. Theoretical Population Biology 61: 349-363.

Wu, R.L., C.X. Ma, and G. Casella. 2007. Linkage analysis with Recombinant Inbred Lines. *In:* Statistical Gen*etics of Quantit*ative Traits. Linkage, Maps, and QTL. Springer (Eds.)

Wricke, G. and Weber, W. E. 1986. Basic concepts of selection. *In*: Quantitative genetics and selection in plant breeding. Walter de Gruyter & Co., Berlin, Germany.

Young, C.T., R.E. Worthington, R.O. Hammons, R.S. Matlock, G.R. Waller, and R.D. Morrison. 1974. Fatty acid composition of spanish peanut oils as influenced by planting locations, soil moisture conditions, variety and season. J. Amer. Oil Chem. Soc. 51: 312-315.

Zeng, Z.B. 1993. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc. Natl. Acad. Sci. 90: 10972-10976.

Zeng, Z.B. 1994. Precision mapping of quantitative trait loci. Genetics 136: 1457-1468.

Chapter II

Heritability of Maturity Level in *Arachis hypogaea* L.: A Study of the Genotypic Variance Components

Franco Villegas Chirinos[1], Thomas G. Isleib*, and Susana R. Milla-Lewis

Formatted for submission to *Peanut Science*

[1] F. Villegas Chirinos, T.G. Isleib, and S.R. Milla-Lewis, Dept. of Crop Science, Box 7620, N.C. State Univ., Raleigh, NC 27695-7620. Received:_____. * Corresponding author email: tom_isleib@ncsu.edu

ABSTRACT

Study of variance components and heritability estimates may provide useful information to define optimal breeding strategies. Two populations, Pop1 with parents PI 313949 (late, large-seeded) and Chico-*ol*$_1$*ol*$_2$ (early, small seeded) and Pop2 with parents PI 365550 (late, large-seeded) and Chico-*ol*$_1$*ol*$_2$, were developed with the main goal of studying inheritance of maturity level in *Arachis hypogaea* L. Ten additional traits: yield (Kg plot^{-1}), traits related to fruit and seed size (length, width, weight), number of seeds per pod, and fruit color (Hunter L, a, and b scores) were also measured and heritability estimates for each trait were obtained. Estimates of additive variance (σ_A^2) for maturity, estimated using an additive-by-additive epistatic model, were negative (and $h^2 = 0$) in both populations. In Pop1, σ_A^2 was also negative for width of 20 fruits, number of seeds in 20 fruits, width of 20 seeds, and pod redness. In Pop2, σ_A^2 was also negative for length, width, and weight of 20 fruits, width of 20 seeds, pod brightness, and pod yellowness. Estimates of additive-by-additive epistatic variance (σ_{AA}^2) were positive in both populations, with the exception of number of seeds in 20 fruits and pod redness for Pop2. These results suggest that, for the populations tested in this study, AxA epistatic gene action influences the traits considerably, and the effects of individual alleles have little or no contribution in the final phenotype. Broad-sense heritabilities (H) for all traits were estimated for families with different levels of inbreeding. Recommendations are given for testing strategies to improve each individual trait separately and for a more realistic case where all traits are to be improved at once.

INTRODUCTION

Early-maturing cultivars of peanut (*Arachis hypogaea* L.) are a requirement in regions of the world with short growing seasons. In temperate zones, late-maturing peanuts often are affected by early frosts or cool night temperatures that can retard the maturation process diminishing yield, grade, and subsequently revenue for farmers (Upadhyaya *et al.*, 2006; Bell *et al.*, 1994). However, peanut maturity level assessment is a complicated task because of the plant's indeterminate flowering and fruiting pattern, the fact that the pods grow below the soil surface, and the quantitative nature of the trait influenced by many genes and the environment (Pattee *et al.*, 1974, Rowland *et al.*, 2006). Several maturity level assessment methods have been developed for peanuts, including indirect methods such as the number of days after planting (DAP) or the heat unit system where weather data is used to predict maturity level and optimum digging date (Sanders *et al.*, 1982a, Rowland *et al.*, 2006). Among the direct methods of evaluation are techniques that use some relative color such as the internal hull color, oil color, methanol extract, the shell-out percentage, and the pod maturity profile or hull-scrape method (William and Drexler, 1981). Other methods use weight and/or weight relationships to assess maturity measuring kernel weight, kernel density, or the seed hull weight ratio (seed/hull ratio maturity index, SHMI). Finally, there are methods based in the correlation of the concentration of a particular compound with maturity stage, for example the arginine maturity index (AMI) and the arachin polypeptide determination (Sanders *et al.*, 1982a; Grimm *et al.*, 1998). The currently preferred method to assess maturity in peanuts is the pod maturity profile or hull-scrape method developed by William and Drexler (1981) (Grimm *et al.*, 1998; Sanders *et al.*, 1982a; Sanders et al 1982b).

In this method, the external layer of the peanut fruit (exocarp) is removed by abrasion or by a high-pressure water stream, revealing the mesocarp which changes color from lighter to darker as the peanut pod matures, progressing from white to yellow, orange, brown and black. A number (100 or 200) of pods are classified into one of four color categories: yellow, orange, brown, or black. A maturity index is then calculated by dividing the number of mature pods (usually the sum of the black and brown classes) by the total number of pods counted after blasting.

In the Virginia-Carolina region, where early maturing cultivars are urgently needed, the commonly grown cultivars belong to the virginia market-type with large fruit, large seeds and late maturity (Chiow and Wynne, 1983). In general, cultivars of the spanish market type are a good source for earliness in crosses with virginia-type parents, but their small fruit and seed sizes and low yield have hindered the development of new early maturing virginia type cultivars with spanish-type parentage. For example, among the preferred sources of earliness we find Chico (PI 565455, GP-2, subsp. *fastigiata* Waldron var. *vulgaris* Harz) (Bailey and Hammons, 1975), a spanish-type germplasm line selected from PI 268661, "APAXUC [Arachis] 370" developed by V.S. Pustovoit at the All-Russia Research Institute of Oil Crops (VNIIMK) in Krasnodar in the former Union of Soviet Socialist Republics (now the Russian Federation). The line was presented to the USDA's National Plant Germplasm System by the Mount Makulu Research Station, Chilanga, Rhodesia (now Zambia) in 1960. Chico is a very early-maturing line with undesirably low biomass and undersized pods and seeds. Nevertheless, this breeding line has been used over a thousand times as a parent for developing early lines at ICRISAT in India (Upadhyaya *et. al.*, 2006). Chico has also been

used in the U.S.A. as a parent in several programs in attempts to shorten maturity. It is a parent of spanish-type cultivars Pronto (Banks and Kirby, 1983) and Spanco (Kirby *et al.*, 1989), a parent of seven registered virginia-type germplasm lines (VGP 2, VGP 3, VGP 4, VGP 5, VGP 6, VGP 10 and VGP 11) developed by the USDA-ARS breeding program at Suffolk, VA (Coffelt and Mozingo, 1998; Coffelt *et. al.*, 1987), and one of 16 parents contributing to the CPES breeding population (Branch and Holbrook, 1991).

Analysis of sources of variability and accurate estimates of heritability are very important for the efficacy of plant breeding methods. Using heritability estimates a breeder is able to assess and compare alternative selection strategies in terms of expected genetic gains. Furthermore, indirect selection methods can be applied to select for highly heritable traits that are correlated with traits with low heritability (Falconer and Mackay, 1996). Calculations of narrow- and/or broad-sense heritabilities, and evaluations of variance components of agronomically important traits, such as yield (Chiow and Wynne, 1983), drought resistance (Songsri *et al*, 2008), peanut flavor (Pattee *et al.*, 1995, Isleib *et al*., 2008) or resistance to diseases (Chiteka *et al*, 1997) have been made in peanuts. Previous reports of heritability estimates and studies of variance components for early maturity in peanuts, using the spanish type breeding line Chico as one of the parents, suggest that early maturity is a fairly highly heritable trait (H>0.5) and that selection for earliness could be performed in early stages of inbreeding (Ali *et al.*, 1994, 1999). Correlations of early maturity with other easily assessed traits as pod size, yield, and oil content suggest that indirect selection strategies could be applied (Chiow and Wynne, 1983; Windham *et al*, 2010).

The goal of this study was to obtain a better understanding of the underlying genetics for early-maturity in peanuts. Estimation of the components of genetic variance would answer critical questions about the trait such as how much of it is controlled by additive or non-additive (epistasis and/or dominance) gene action. Estimation of heritability (h^2 and H) will allow the identification of optimal breeding strategies for the specific population under observation.

MATERIALS AND METHODS

Plant material and population development. Two populations of recombinant inbred lines (RILs) were developed, one with 132 F_6-derived families derived from the cross of plant introduction PI 313949 and a high-oleic backcross derivative of PI 565455 (Chico). This high-oleic line was designated "Chico-ol_1ol_2." The second population comprised 50 F_6-derived families developed from the cross of PI 365550 with Chico-ol_1ol_2. PI 565455 is a spanish-type (*A. hypogaea* subsp. *fastigiata* Waldron var. *vulgaris* Harz) germplasm line selected from PI 268661, presented to the National Plant Germplasm System by the Mount Makulu Research Station, Chilanga, Rhodesia (now Zambia) in 1960 (Bailey and Hammons, 1975). The line was sent to Rhodesia by V.S. Pustovoit from the All-Russia Research Institute of Oil Crops (VNIIMK) in Krasnodar in the former Union of Soviet Socialist Republics (now the Russian Federation) where it had the designation "APAXUC [*Arachis*] 370". Chico is a very early maturing cultivar (20 to 30 days earlier than the earliest commercial cultivars) with very small tan (single colored) seeds (USDA GRIN, 2011a). PI 313949, is a Bolivian overo-type peanut (*A. hypogaea* subsp. *hypogaea* var. *hypogaea*) with local name "Overo Chiquitano,". This line has pronounced late maturity (130 DAP), it is normal oleic, and it has large pods and large variegated (red and white) seeds, whose average weight is more than twice that of seeds of Chico (USDA GRIN, 2011b). PI 365550, another overo-type (*A. hypogaea* subsp. *hypogaea* var. *hypogaea*) Bolivian introduction with local name "Mani Gris," is also very late maturing with variegated (red and white) seeds that are almost three times the average weight of Chico (USDA GRIN, 2011c).

Crosses were made in the summer of 2000 at the NCSU campus. In 2001, the F_1 was grown at the N.C. Dept. of Agriculture and Consumer Services Peanut Belt Research Station (PBRS) near Lewiston, NC, using standard agronomic practices (weed and pest control, plant density, etc). In 2002, the F_2 was grown at the same location and with similar practices, but individual F_2 plants were harvested producing F_2-derived families. These families were advanced until the F_6 generation using the modified pedigree method (single seed descent). In 2007, two individual F_6 plants were chosen at random from each F_2-derived family, forming the F_6-derived families evaluated in this study. In 2008 and 2009, the two populations were planted at PBRS (one location, with two replications per year) in a simple 14x14 square lattice designs including the 182 F_6-derived families, the three parents, and 11 checks including: NC 7, Gregory, Perry, Phillips, Brantley, PI 371853 / 2*N90010E, Gregory / N91040, NC 12C*2 / N96076L, Florunner, and Georgia Green (some varieties with known maturity level and agronomic traits). Each plot consisted of two rows with 3.7 m in length, with rows spaced 91 cm apart and seeding rates of 4 seeds per meter. Plots were planted, irrigated and treated conventionally for pests and weed control. The replicated families were dug 125 and 130 DAP in 2008 and 2009, respectively. In both years, a stationary thresher was used at harvest. A single bag of about 1 kg of pods was collected from each individual plot and used for all the phenotypic evaluations carried out in this study.

Phenotypic Evaluation. Maturity level was measured by the hull-scrape method, colloquially known as "pod-blasting" (Williams and Drexler, 1981). The exocarps were removed from a sample of 100 pods using a Task Force® (Rexon Industrial Corp, Taiwan, ROC) 2000 PSI electric high-pressure washer with 2000 maximum PSI, 1.6 GPM maximum

flow rate, and a turbo spray nozzle with an orifice size of 1.5 mm. Pods were separated into categories reflecting differential maturity based on mesocarp color: white, yellow, orange, brown or black. The pods in the five color categories were counted, and a maturity index was calculated as the percentage of pods falling in the brown and black categories. Four other quantitative traits were measured: pod color and brightness, plot yield, peanut pod dimensions, and peanut seed dimensions. Pod color and brightness were measured using a Hunterlab D25-PC2 colorimeter in a procedure similar to the one described by Isleib *et al.* (1997). Yield was estimated by weighing all the pods produced by the whole plot of each one of the F_6-derived families harvested per year (2008 and 2009). Peanut pod and seed dimensions were determined by measuring with a ruler the length and width of 20 pods and 20 seeds of each of the F_6-derived families. Pods and seeds were laid end-to-end or side-by-side and total length or width was recorded. All phenotypic evaluations were carried out for two years, 2008 and 2009.

Data Analysis. Phenotypic data for the two years were analyzed using the mixed model (PROC MIXED) and the general linear model (PROC GLM) procedures of SAS statistical software Version 9.2 (SAS Inst. Inc., 2008). For the study of variance components, variation among the lines tested in the field trials was partitioned into portions reflecting the difference between the means of the experimental populations (including parents) and the check cultivars, variation among the check cultivars, the contrast between the parents of each population, and the contrast between the mean of the hybrid population and the mean of its parents (a test for heterosis or average non-additive genetic effects in the population). These were considered to be fixed effects. The remaining variation was partitioned to reflect

portions due to the F_2-derived families within each population (a measure of half the additive genetic variance within the population), and variation among the F_6-derived families within F_2-derived families (a measure of the remaining additive genetic variance and most of any non-additive genetic variance present). These were considered to be random effects and variance components were estimated using an iterative maximum likelihood method (SAS Inst. Inc., 2008). Coefficients for the expected additive and non-additive components of genetic variance among and within F_2-derived families were calculated using the method of Cockerham (1983). Components of genetic variance were estimated by equating the observed mean squares to their expectations under the genetic model and solving for the unknown variance components. An analogous partition of the year-by-genotype sum of squares was made, and year-by-additive and year-by-nonadditive variance components were estimated. Predicted values for narrow-sense heritabilites were computed for F_2 plants and for families at different levels of inbreeding and extents of testing.

RESULTS AND DISCUSSION

Maturity level. The contrasts between the mean maturity indices of the female and male parents used in Pop1 (PI 313949 / Chico-*ol*$_1$*ol*$_2$) and Pop2 (PI 365550 / Chico-*ol*$_1$*ol*$_2$) were highly significant ($P<0.0001$ for Pop1 and Pop2) (Table 1) with Chico-*ol*$_1$*ol*$_2$ being earlier than either of the two overo-type parents. The means of the F_6-derived families from Pop1 and Pop2 were found not to differ from the respective midparental means ($P=0.5745$ for Pop1 and $P=0.0780$ for Pop2), a preliminary indicator of the absence of a high average degree of dominance in either population. The check cultivars' mean maturity index

(0.5992), was significantly higher than the mean maturity of either of the overo-type parents, with P-values of 0.0162 for PI 313949 (mean maturity index=0.4298), and <0.0001 for PI 365550 (mean maturity index=0.1752). The cultivars' mean was significantly lower than the mean for Chico-ol_1ol_2 (P-value=<0.0001). The cultivars' mean value was also found to be significantly lower from the mean value for either Pop1 or Pop2 (P-values= 0.0132 and 0.0363 respectively). The populations developed for these study were earlier than the average for the commercial cultivars. However, the F_6-derived lines that make each population showed undesirable agronomic traits (small fruit and seed dimensions, dark hull color).

Other traits. The mean value for Pop2 was significantly different from its midparent-value for length, width, and weight of 20 fruits and of 20 seeds, suggesting that dominance or some other non-additive form of gene action was expressed by the genes influencing the traits. Because the population is relatively highly inbred (F=31/32 for the S_5 considering the F_2 population to be the S_0), one would not expect much heterozygosity to be left. Therefore there is little opportunity for the expression of dominance effects in the population. PI 365550 has higher values than the male line Chico-ol_1ol_2 for all the mentioned traits, *i.e.*, it has bigger pods and seeds. For all these traits, the mean values for Pop2 were not significantly different from the mean value of PI 365550, the female parental line used in the cross. These results, combined with the fact that peanut is an allotetraploid species for which duplicated genes should occur frequently, raise duplicate gene action as a possible genetic mechanism controlling the mentioned traits. Duplicate gene action would allow for expression of substantial additive-by-additive epistatic effects in inbred generations.

Similar results are observed in Pop1 for the traits width and weight of 20 fruits and length, width, and weight of 20 seeds where the means of the $F_{6:7}$ families were significantly different from the mid-parent values for that population. The $F_{6:7}$ family mean values for Pop1 were not significantly different from the mean values of PI 313949, the female parental line. PI 313949 had higher values than the male line Chico-*ol*$_1$*ol*$_2$ for all the mentioned traits.

These findings suggest a high degree of non-additive gene action controlling the mentioned traits. Again, it is unlikely that the observed relationships among the parental and family means could be ascribed to dominance effects in a highly inbred population. Another more complex gene action mechanism, epistasis involving two or more loci, would likely be involved. Again because of the rare occurrence of heterozygosity in F_6-derived families, additive-by-dominance and dominance-by-dominance forms of epistatic interactions should not be often expressed. One would expect additive-by-additive forms to prevail in a highly inbred population. Similar results were observed for six traits (five in Pop2), and finding the same mechanism producing extreme phenotypes in different traits would be improbable. However, all these traits relate to pod and seed dimensions, and it is possible that many of the controlling genes are shared among them. Finally, another explanation may be the presence of strong selection favoring the female genotypes during population development. The female lines share many agronomic characteristics, including pod and seed dimension traits, because they belong to the same subspecies (subsp. *hypogaea*) and were collected in the same region. In a parallel study of these two populations (Villegas *et al*., 2010), we carried out a screening of PI 313949, PI 365550 and Chico-*ol*$_1$*ol*$_2$ using 453 simple sequence repeats (SSR) markers. A total of 330 (72.8%) of these SSR markers were not polymorphic between

PI 313949 and PI 365550. These results suggest a high degree of similarity, in this case at the molecular marker level, between these genotypes.

It was not possible to simultaneously estimate more than two types of genetic variances because there were only two different family types in the mating design used to produce the experimental populations, *i.e.*, $F_{2:7}$ ($S_{0:5}$) families and $F_{6:7}$ ($S_{4:5}$) families within them. While it would be useful to have simultaneous estimates of additive, (σ_A^2), dominance (σ_D^2), and epistatic (σ_{AA}^2) variances, it was possible only to estimate two simultaneously. Therefore, two alternative genetic models were considered: an additive-dominance (A-D) model and an additive with additive-by-additive (A-AA) epistatic model. For both Pop1 and Pop2, estimates of additive variance were very small using the additive-dominance genetic model and negative for the additive-AA epistatic model (Table 2). The non-additive component was much higher than the additive component in both populations and both models. Given that there was very little evidence of dominance based on the comparison of the mean of F_6-derived families with the midparent mean, it is most reasonable to give more credence to the A-AA model. A similar partition of variance components has been reported in peanuts for sensory descriptors on another population (Isleib *et al.*, 2003) and possible explanations were given. Mainly, the very small coefficients on σ_D^2 in the covariances of inbred relatives may have inflated the estimates of σ_D^2, but the coefficients on σ_{AA}^2 were greater than or equal to those on σ_A^2 in the relevant covariances. Using the dominance model produced positive estimates of narrow sense heritability (h^2). The estimate of the additive variance component in the A-AA model was negative. The model that better fits the allotetraploid self-pollinated

peanut genetics is the A-AA epistatic model. For that reason, we have to conclude that σ^2_A and h^2 for maturity level is zero. Nevertheless, estimation of broad-sense heritability (H) is still useful because the additive-by-additive effects can be fixed in pure lines.

From the estimates of broad-sense heritability for maturity level at different stages of inbreeding (Table 2), we can infer the best resource allocation for improving maturity in Pop1 and Pop2. In Pop1, inbreeding beyond the F_3 appears to be ineffective as a way to increase H. One extra generation of inbreeding beyond the F_3 produces only a 7% increase in H. For Pop 2, going further than four generations of inbreeding (F_4) produces an increase of only 5% in H. In general, it would be advisable to wait until the F_4 and select individual plants producing $F_{4:5}$ lines and testing these for the best combination of years (y) and repetitions (r) to achieve high H with fewer resources. Considering that it is cheaper (and obviously faster) to test over more reps in fewer years than fewer reps in more years, testing $F_{4:5}$ lines for two years with three replications per year seems to be a reasonable testing program to improve maturity.

Variance components and heritability estimates were obtained, using the additive-epistasis model for yield, pod brightness (Hunter L score), pod redness (Hunter a score), pod yellowness (Hunter b score), 20 pods and 20 seeds length and width, 20 pods and 20 seeds weight, and number of seeds per 20 pods (Table 3). Additive variance (σ^2_A) estimates for maturity, width of 20 fruits, and width of 20 seeds were negative for both populations. These results are contradictory with the ones reported by Ali and Wynne (1994), where h^2 for individual F_2 plants obtained by parent-offspring regression for maturity index were 0.42 and 0.44 for two populations with Chico and a virginia type cultivar (NC 7 or 70-30) as parents.

However, our results are partially in agreement with Ali *et al.* (1999), where negative additive variance for maturity was reported in one cross of a var. *hypogaea* (No. 334) with a spanish type cultivar (ICGSE-4) as the source of earliness. We should mention that in the same study by Ali (1999), an estimate h^2=0.87 was obtained for maturity in a second population with the same spanish-type parent (ICGSE-4) but a different var. *hypogaea* parent, NC 9 (Wynne *et al.*, 1986). A previous study by Chiow and Wynne (1983) also reported negative total genetic variance for maturity in a cross involving a virginia-type cultivar (NC 6) and a spanish-type line (922). In Pop1, σ^2_A estimates were also negative for number of seeds in 20 fruits, and pod redness. In Pop2, σ^2_A estimates were also negative for: yield, length of 20 fruits, weight of 20 fruits, pod brightness, and pod yellowness. Additive-by-additive variance estimates (σ^2_{AA}) were positive in both populations, with the exception of number of seeds in 20 fruits and pod redness for Pop2. Most h^2 estimates were zero in Pop2, with only the estimate for number of seeds in 20 fruits (0.24) being larger than 0.05. The other estimates of σ^2_{AA} with positive values for Pop2 were for length of 20 seeds (0.04), weight of 20 seeds (0.01), and pod redness (0.05). These results suggest that, for the populations tested in this study, additive-by-additive epistatic gene action influences the traits considerably, and the additive genetic effects of individual alleles have little or no influence on the final phenotype. In Pop1, h^2 estimates were higher than 0.05 only for length of 20 fruits (0.27), weight of 20 fruits (0.25), and length of 20 seeds (0.19). Comparable results were reported by Chiow and Wynne (1983) for length of 20 fruits, and weight of 20 fruits; with h^2 estimates of 0.16 and 0.17, respectively, when estimated from parent offspring

regression using F_2 and F_3 generations. However, when regression using the F_5 and F_6 (grown in different years) was employed, the heritability estimate for weight of 20 fruits increased to 0.50. Narrow sense heritabilities for pod length (0.28-0.57), number of seeds in 30 pods (0.02-0.38), and weight of 100 seeds (0.50-0.71) were reported by Ali and Wynne (1994). Heritability estimates for yield per plant (0.07-0.32), length of 20 pods (0.43-0.91), number of seeds in 50 pods (0.52-0.83), and weight of 100 seeds (0.23-0.62) have been reported by Ali *et al.* (1999). Yield per plot of peanuts has been reported as a trait with very low narrow-sense (h^2<0.1) heritability estimates (Ali *et al.,* 1999). Apart from the hull color parameters, yield is the only trait among the ones we tested with consistently very low H for F_2 plants (<0.06) across populations (Pop1 and Pop2) (Table 3). In general, for most if not all crops, yield is considered a trait with very low heritability because it is influenced by many non-genetic factors. Our results agree with this premise, and our recommendation for the testing program designed to improve yield would be to advance the populations to F_4- or F_5-derived families and start a testing program with at least three years and two replications per year (Table 4).

Narrow sense heritability (h^2) estimates for pod length and pod weight were zero for Pop2 and very similar in Pop1 (0.27 and 0.25, for F_2 plants), indicating that some progress is expected to be made in evaluating F_2-derived lines from Pop1 but they should be evaluated for many years and repetitions. For pod length and weight h^2 is 0.58 and 0.55, respectively when $F_{2:3}$ lines are evaluated in three years and three reps.

Broad sense heritability (H) for F_2 plants was similar among traits related to fruit dimensions: length of 20 fruits, width of 20 fruits, and weight of 20 fruits. They ranged

from 0.38 to 0.40 in Pop1, and 0.40 to 0.53 in Pop2. Therefore, early generation testing appears to be an option for improvement of fruit dimension related traits. For length of 20 fruits, going from the F_2 to the F_3 before generating lines (single plant selection) produces an increase in H of only 7% when three years with three replications per year are used, and of 10% when two years and three reps are used (Table 5). Testing $F_{2:3}$ lines for two years with three reps per year achieves an H of 0.7939 in Pop1 and of 0.8319 in Pop2. Both values are relatively close to the value for $F_{6:7}$ lines tested for 3 years with 3 reps, 0.9359 in Pop1 and 0.9654 in Pop2. An appropriate testing scheme for fruit dimension traits would be to make plant selections in the F_2 or F_3 and test for two years with three repetitions per year.

From the contrasts estimated above, number of seeds in 20 pods showed no significant difference between the parental lines involved in Pop2. The numerical value for the average number of seeds in 20 pods for the parental lines was very close (around 35). The averages for Pop2 and PI 365550 were also not different at any significance level of Type I error. These results indicate that there is not very much variability in Pop2 for number of seeds. Number of seeds in 20 pods is one of the two traits in Pop2 (the other is pod redness) with estimated negative non-additive variance. Number of seeds in 20 pods has zero narrow sense heritability and positive non-additive variance in Pop1. Estimates of H for individual F_2 plants for Pop1 and Pop2 were 0.20 and 0.25, these relatively high H estimates are unexpected because of the lack of parental variability for this trait. The only source of genetic variance in Pop1 seems to be non-additive, whereas in Pop2 additive variance seems to be the only component of the genetic variance (Table 3). The disparity in results between

populations make difficult to make any general recommendation for testing programs for number of seeds in 20 pods.

For length of 20 seeds, the estimate of narrow sense heritability for individual F_2 plants was much higher in Pop1 (0.1916) than in Pop2 (0.0368), indicative of a higher influence of additive gene action in Pop1. However, not much improvement is expected even in Pop1 when testing F_2-derived lines for three years with three reps per year (h^2=0.3542). From the H estimates (Table 6), we can see that testing $F_{3:4}$ lines for two years and two reps gives an H estimate of 0.8559 in Pop1 and 0.8469 in Pop2 whereas testing $F_{6:7}$ lines for three years and three reps produces an H of 0.9486 and 0.9509 for Pop1 and Pop2, respectively. The reduction in H due to using $F_{3:4}$ instead of $F_{6:7}$ lines and testing for one less year and repetition seem to be overcome by the savings in time and resources in population development and testing. Resource savings will allow the breeder to test more lines and, because a high H is still reached, the probability of success is increased.

For width and weight of 20 seeds, additive variance (and consequently h^2) was estimated to be very small or zero. H estimates for individual F_2 plants were higher for weight of 20 seeds (0.3355 and 0.3092, for Pop1 and Pop2) than for width of 20 seeds (0.1942 and 0.1736, for Pop1 and Pop2, Table 3). However, both traits behave similarly in the change of H across inbreeding stages and combinations of years and reps, and the same conclusions are reached when searching for an optimal testing strategy. For weight of 20 seeds, testing $F_{3:4}$ lines for two years and three reps per year get an H of 0.8437 and 0.8197 (Pop1 and Pop2). Testing $F_{6:7}$ lines for three years and three reps produces an H of 0.9298 and 0.9190 (Table 7). Again, resource savings tops gain in H and our recommendation

would be testing $F_{3:4}$ lines for two years and three reps. However, a higher response to selection would be expected for weight than for width of 20 seeds.

The traits related with pod color were pod brightness, redness and yellowness. All of them exhibited very small genetic variances, both additive and non-additive (Table 3). In the contrasts among parents and population means, none of these traits showed a significant difference between the parents used in Pop 1, and only pod brightness showed significance difference between parents in Pop2. The estimates for h^2 and H for the F_2 for these traits were always low, less than 6% in all traits and populations with the exception of pod brightness in Pop1 where H was 0.0859. Therefore, we must conclude that these crosses are not a good option to assess h^2 and H estimates for pod color related traits and recommendations for testing strategies cannot be made.

Considering a more realistic scenario, in a breeding program one would not conduct separate testing programs for separate traits. It is more reasonable to conduct a single testing program using the same lines and measuring all the traits of interest. In that case, one must test the appropriate families extensively enough to achieve high enough H for the traits that will be the most difficult to improve, *i.e.*, the traits with the lowest H. Leaving aside the traits related with fruit color (Hunter L-, a-, and b-scores) for the reasons discussed above, these traits are yield and maturity (Table 3). As mentioned before, these traits should be evaluated for families developed after reaching at least the third generation of inbreeding, *i.e.* using F_4-derived lines. We would recommend a testing program of at least three years with three repetitions per year, giving an H of 0.3764 for yield and 0.8500 for maturity in Pop1; and 0.4808 for yield and 0.7104 for maturity in Pop2 (Tables 2 and 4). The breeder would

have the option of testing the other traits for only two years or in fewer than three replicates per year. This reduced effort on measuring traits with higher heritabilities would be especially efficient if the phenotyping of many lines is very expensive or time consuming, as it may be the case with maturity level, and measuring only yield for more than two years is a valid option. Further increase in H for yield is possible when testing for more than three years, and the extent of testing for this important trait will depend on the program resources.

REFERENCES

Ali, N., and J.C. Wynne. 1994. Heritability estimates and correlation studies of early maturity and other agronomic traits in two crosses of peanuts (*Arachis hypogaea* L.). Pak. J. Bot., 26(1): 75-82.

Ali, N., J.C. Wynne, and J.P. Murphy. 1999. Estimation of genetic effects and heritability for early maturity and agronomic traits in peanuts (*Arachis hypogaea* L.). Pak. J. Bot., 31(2): 323-335.

Bailey, W.K., and R.O. Hammons. 1975. Registration of Chico peanut germplasm. Crop Sci. 15: 105.

Banks, D.J., and J.S. Kirby. 1983. Registration of Pronto peanut (Reg. No. 28). Crop Sci. 23: 184.

Bell, M. J., R. C. Roy, M. Tollenaar, and T. E. Michaels. 1994. Importance of variation in chilling tolerance for peanut genotypic adaptation to cool, short-season environments. Crop Sci. 34: 1030-1039.

Branch, W.D., and C.C. Holbrook. 1991. Registration of CPES peanut germplasm population. Crop Sci. 31: 497-498.

Chiow, H.Y., and J.C. Wynne. 1983. Heritabilities and genetic correlations for yield and quality traits of advanced generations in a cross of peanut. Peanut Sci. 10: 13-17.

Chiteka, Z.A., D.W. Gorbet, F.M. Shokes and T.A. Kucharek. 1997. Components of resistance to early leaf spot in peanut-genetic variability and heritability. Soil Crop Sci. Soc. Fl. Proc., 56: 63-68.

Cockerham, C. 1983. Covariances of relatives from self-fertilization. Crop Sci., 23: 1177-1180.

Coffelt, T.A., D.M. Porter, J.C. Smith, and R.W. Mozingo. 1987. Registration of six peanut germplasm lines with multiple resistance. Crop Sci. 27:1 319.

Coffelt, T.A., and R.W. Mozingo. 1998. Registration of VGP 10 peanut germplasm. Crop Sci. 38: 573-574.

Coffelt, T.A., R.W. Mozingo, and D.A. Herbert, Jr. 1998. Registration of VGP 11 peanut germplasm. Crop Sci. 38: 1412.

Falconer, D.S., and T.F. Mackay. 1996. Continuous variation: metric characters, p. 100-104. *In* Longman A.W. (Ed.), Quantitative Genetics. Pearson Education Limited, Essex, UK.

Grimm, C.C., E.T. Champagne, and T.H. Sanders. 1998. Determination of peanut maturity using a hunter colorimeter. Peanut Sci. 25: 99-103.

Isleib, T.G., H.E. Pattee, and P.W. Rice. 1997. A laboratory colorimeter method to measure pod brightness in virginia-type peanuts. Peanut Sci. 24: 81-84.

Isleib, T.G., H.E. Pattee, and F.G. Giesbrecht. 2003. Narrow-sense heritability of selected sensory descriptors in virginia-type peanuts (*Arachis hypogaea* L.) Peanut Sci. 30: 64–66.

Isleib, T.G., B.L. Tillman, H.E. Pattee, T.H. Sanders, K.W. Hendrix, and L.O. Dean. 2008. Genotype-by-environment interactions for flavor attributes of breeding lines in the Uniform Peanut Performance Test. Peanut Sci. 35: 55-60.

Kirby, J.S., D.J. Banks, and J.R. Sholar. 1989. Registration of 'Spanco' peanut. Crop Sci. 29: 1573-1574.

Pattee, H.E., E.B. Johns, J.A. Singleton, and T.H. Sanders. 1974. Composition changes of peanut fruit parts during maturation. Peanut Sci. 1: 57–62.

Pattee, H.E., F.G. Giesbrecht, and T.G. Isleib. 1995. Roasted peanut flavor intensity variations among U.S. peanut genotypes. Peanut Sci. 22: 158-162.

Rowland, D.L., R.B. Sorensen, C.L. Butts, and W.H. Faircloth. 2006. Determination of maturity and degree day indices and their success in predicting peanut maturity. Peanut Sci. 33: 125-136.

SAS Inst. Inc. 2008. SAS/STAT 9.2 User's Guide. SAS Inst., Inc. Cary, NC.

Sanders, T.H., A.M. Schubert, and H.E. Pattee. 1982a. Maturity methodology and postharvest physiology, p. 624-654. *In* H.E. Pattee and C.T. Young (eds.) Peanut Science and Technology. Amer. Peanut Res. Educ. Soc., Inc., Yoakum, Texas, USA.

Sanders, T.H., J.A. Lansden, R.L. Greene, J.S. Drexler, and E.J. Williams. 1982b. Oil characteristics of peanut fruit separated by a non-destructive maturity classification method. Peanut Sci. 9: 20-23.

Songsri, P., S. Jogloy, T. Kesmala, N. Vorasoot, C. Akkasaeng, A. Patanothai, and C.C. Holbrook. 2008. Traits and correlation of drought resistance and agronomic traits in peanut. Crop Sci. 48: 2245-2253.

Upadhyaya, H.D., L.J. Reddy, C.L.L. Gowda, and S. Singh. 2006. Identification of diverse groundnut germplasm:Sources of early maturity in a core collection. Field Crops Res. 97: 261-271.

USDA-ARS National Genetic Resources Program. 2011a. Germplasm Resources

Information Network (GRIN) database. Observations for accession PI 565455. Available at http://www.ars-grin.gov/cgi-bin/npgs/acc/obs.pl?1460426. National Germplasm Resources Laboratory, Beltsville, MD.

USDA-ARS National Genetic Resources Program. 2011b. Germplasm Resources Information Network (GRIN) database. Observations for accession PI 313949. Available at http://www.ars-grin.gov/cgi-bin/npgs/acc/obs.pl?1234443. National Germplasm Resources Laboratory, Beltsville, MD.

USDA-ARS National Genetic Resources Program. 2011c. Germplasm Resources Information Network (GRIN) database. Observations for accession PI 365550. Available at http://www.ars-grin.gov/cgi-bin/npgs/acc/obs.pl?1273705. National Germplasm Resources Laboratory, Beltsville, MD.

Villegas Chirinos, F., T.G. Isleib, S.J. Knapp, and S.R. Milla-Lewis. 2010. Molecular mapping of QTL for early maturity in peanut. Poster presentation at the 2010 ASA International annual meeting. Oct. 31-Nov. 4, Long Beach, CA . http://a-c-s.confex.com/crops/2010am/webprogram/Paper60561.html

Williams, E. J., and J. S. Drexler. 1981. A non-destructive method for determining peanut pod maturity. Peanut Sci. 8: 134-141.

Windham, W.R., C.V.K. Kandala, J. Sundaram, and R.C. Nuti. 2010. Determination of peanut pod maturity by near-infrared reflectance spectroscopy. Transactions of the ASABE, 53(2): 491-495.

Wynne J.C., R.W. Mozingo, and D.A. Emery. 1986. Registration of ‘NC 9’ peanut. Crop Sci. 26: 197.

Table 1. Probability values for relevant contrasts for all traits considered.

	Population 1			Population 2		
	Female vs. male parent	$F_{6:7}$ mean vs. midparent	$F_{6:7}$ mean vs. female parent	Female vs. male parent	$F_{6:7}$ mean vs. midparent	$F_{6:7}$ mean vs. female parent
Maturity (% brown and black pods)	<0.0001	0.5745	0.0007	<0.0001	0.0780	<0.0001
Yield (kg ha^{-1})	0.0002	0.3173	0.0010	0.0024	0.9811	0.0406
Length of 20 fruits (cm)	0.0012	0.2121	0.1840	0.0713	0.0002	0.1088
Width of 20 fruits (cm)	<0.0001	<0.0001	0.8539	<0.0001	<0.0001	0.2962
Weight of 20 fruits (g)	<0.0001	0.0268	0.0686	<0.0001	<0.0001	0.8100
No of seeds in 20 fruits	0.0470	0.5358	0.0708	0.8314	0.2462	0.4754
Length of 20 seeds (cm)	0.0001	0.0002	0.9392	<0.0001	<0.0001	0.6541
Width of 20 seeds (cm)	0.0001	0.0164	0.3543	0.0010	<0.0001	0.3857
Weight of 20 seeds (g)	<0.0001	0.0019	0.4225	<0.0001	<0.0001	0.4703
Pod brightness (Hunter L score)	0.3058	0.2668	0.9090	0.3302	0.7354	0.6834
Pod redness (Hunter a score)	0.3127	0.8558	0.5781	0.0046	0.5079	0.0162
Pod yellowness (Hunter b score)	0.8349	0.2287	0.3064	0.7778	0.4601	0.4619

Table 2. Estimates of broad sense heritability (H) for maturity level at different stages of inbreeding and combinations of years (y) and repetitions per year (r).

		Population 1					Population 2				
Years (y)	Reps in years (r)	$F_{2:3}$	$F_{3:4}$	$F_{4:5}$	$F_{5:6}$	$F_{6:7}$	$F_{2:3}$	$F_{3:4}$	$F_{4:5}$	$F_{5:6}$	$F_{6:7}$
1	1	0.1706	0.3163	0.3864	0.4196	0.4357	0.0997	0.1995	0.2532	0.2802	0.2936
1	2	0.2914	0.4806	0.5575	0.5912	0.6069	0.1648	0.3075	0.3767	0.4096	0.4256
1	3	0.3816	0.5813	0.6539	0.6844	0.6984	0.2107	0.3753	0.4498	0.4842	0.5006
2	1	0.2914	0.4806	0.5575	0.5912	0.6069	0.1813	0.3326	0.4041	0.4378	0.4540
2	2	0.4513	0.6492	0.7159	0.7431	0.7554	0.2830	0.4704	0.5473	0.5812	0.5971
2	3	0.5524	0.7352	0.7908	0.8127	0.8224	0.3481	0.5457	0.6205	0.6524	0.6672
3	1	0.3816	0.5813	0.6539	0.6844	0.6984	0.2494	0.4277	0.5043	0.5387	0.5550
3	2	0.5524	0.7352	0.7908	0.8127	0.8224	0.3719	0.5712	0.6445	0.6755	0.6897
3	3	0.6492	0.8064	0.8500	0.8668	0.8742	0.4447	0.6431	0.7104	0.7379	0.7504

Table 3. Estimates of variance components, narrow sense (h^2) and broad sense heritability (H) in individual F_2 plants for Pop1 and Pop2.

Population 1 (PI 313949 / Chico-*ol1ol2*)

	$\hat{\sigma}^2_Y$	$\hat{\sigma}^2_{R(Y)}$	$\hat{\sigma}^2_A$	$\hat{\sigma}^2_{AA}$	$\hat{\sigma}^2_{YA}$	$\hat{\sigma}^2_{YAA}$	$\hat{\sigma}^2$	h^2	H
Maturity(% black and brown pods)	0.0000	0.0025	-0.0030	0.0039	0.0000	0.0000	0.0164	0.0000	0.1706
Yield (kg plot^{-1})	0.4658	0.8627	0.0757	0.0112	0.0000	0.0000	0.2250	0.0461	0.0530
Length of 20 fruits (cm)	0.0000	1.3604	12.6735	6.1002	0.5527	0.2124	25.5859	0.2726	0.4039
Width of 20 fruits (cm)	0.0000	0.0094	-0.1066	1.5782	-0.1470	0.1470	2.6688	0.0000	0.3584
Weight of 20 fruits (g)	0.0000	0.1645	8.9337	4.5636	-2.0594	2.0594	19.4655	0.2539	0.3836
No seeds in 20 fruits	0.0000	0.2721	-0.0697	1.9929	0.0000	0.0000	7.6640	0.0000	0.2007
Length of 20 seeds (cm)	0.0649	0.3962	1.5373	2.1953	0.5240	-0.1399	3.3071	0.1916	0.4651
Width of 20 seeds (cm)	0.0000	0.0218	-0.7588	0.7588	-0.0267	0.0267	3.0995	0.0000	0.1942
Weight of 20 seeds (g)	0.0000	0.0269	0.1811	1.3213	0.3158	-0.0585	2.6326	0.0404	0.3355
Pod brightness (Hunter L score)	9.2086	1.3952	0.8775	0.5131	-0.2622	0.2622	3.9375	0.0542	0.0859
Pod redness (Hunter a score)	0.0000	0.0559	-0.0076	0.0076	-0.0001	0.0064	0.1338	0.0000	0.0375
Pod yellowness (Hunter b score)	0.8613	0.6226	0.0088	0.1274	0.0457	-0.0155	0.6694	0.0038	0.0583

Population 2 (PI 365550 / Chico-*ol1ol2*)

	$\hat{\sigma}^2_Y$	$\hat{\sigma}^2_{R(Y)}$	$\hat{\sigma}^2_A$	$\hat{\sigma}^2_{AA}$	$\hat{\sigma}^2_{YA}$	$\hat{\sigma}^2_{YAA}$	$\hat{\sigma}^2$	h^2	H
Maturity(% black and brown pods)	0.0000	0.0025	-0.0004	0.0024	0.0026	-0.0009	0.0165	0.0000	0.0997
Yield (kg plot^{-1})	0.4658	0.8627	-0.0856	0.0856	-0.0218	0.0218	0.2209	0.0000	0.0516
Length of 20 fruits (cm)	0.0000	1.3604	-21.4201	21.5472	0.0000	0.0000	24.7657	0.0000	0.4520
Width of 20 fruits (cm)	0.0000	0.0094	-2.0332	3.1980	0.3002	-0.1022	2.5313	0.0000	0.5296
Weight of 20 fruits (g)	0.0000	0.1645	-6.3035	13.0451	0.0000	0.0000	18.9064	0.0000	0.4062
No seeds in 20 fruits	0.0000	0.2721	2.5775	-0.1905	0.0000	0.0000	7.8478	0.2409	0.2409
Length of 20 seeds (cm)	0.0649	0.3962	0.2434	2.4166	-0.1750	0.1750	3.3205	0.0368	0.4020
Width of 20 seeds (cm)	0.0000	0.0218	-0.5564	0.6571	0.0000	0.0000	3.1056	0.0000	0.1736
Weight of 20 seeds (g)	0.0000	0.0269	0.0348	1.3794	-0.5022	0.5022	2.6300	0.0076	0.3092
Pod brightness (Hunter L score)	9.2086	1.3952	-0.6185	0.8947	0.6873	-0.2340	3.9377	0.0000	0.0555
Pod redness (Hunter a score)	0.0000	0.0559	0.0100	-0.0024	0.0000	0.0000	0.1345	0.0499	0.0499
Pod yellowness (Hunter b score)	0.8613	0.6226	-0.1501	0.1501	0.1417	-0.0310	0.6716	0.0000	0.0613

Table 4. Estimates of broad sense heritability (H) for yield per plot (kg) at different stages of inbreeding and combinations of years (y) and repetitions per year (r).

		Population 1					Population 2				
Years (y)	Reps in years (r)	$F_{2:3}$	$F_{3:4}$	$F_{4:5}$	$F_{5:6}$	$F_{6:7}$	$F_{2:3}$	$F_{3:4}$	$F_{4:5}$	$F_{5:6}$	$F_{6:7}$
1	1	0.0530	0.0819	0.0969	0.1044	0.1082	0.0516	0.1091	0.1429	0.1607	0.1697
1	2	0.0792	0.1208	0.1417	0.1521	0.1574	0.0767	0.1575	0.2029	0.2261	0.2378
1	3	0.0949	0.1434	0.1675	0.1795	0.1854	0.0916	0.1849	0.2359	0.2616	0.2745
2	1	0.1006	0.1515	0.1767	0.1891	0.1953	0.0982	0.1968	0.2501	0.2769	0.2902
2	2	0.1468	0.2155	0.2482	0.2641	0.2719	0.1425	0.2722	0.3373	0.3688	0.3842
2	3	0.1734	0.2508	0.2869	0.3043	0.3128	0.1678	0.3120	0.3817	0.4148	0.4308
3	1	0.1436	0.2112	0.2435	0.2592	0.2669	0.1404	0.2688	0.3335	0.3648	0.3801
3	2	0.2051	0.2918	0.3312	0.3499	0.3591	0.1996	0.3594	0.4330	0.4671	0.4835
3	3	0.2393	0.3343	0.3764	0.3962	0.4058	0.2322	0.4049	0.4808	0.5153	0.5316

Table 5. Estimates of broad sense heritability (H) for length of 20 fruits (cm) at different stages of inbreeding and combinations of years (y) and repetitions per year (r).

Years (y)	Reps in years (r)	Population 1					Population 2				
		$F_{2:3}$	$F_{3:4}$	$F_{4:5}$	$F_{5:6}$	$F_{6:7}$	$F_{2:3}$	$F_{3:4}$	$F_{4:5}$	$F_{5:6}$	$F_{6:7}$
1	1	0.4039	0.5416	0.5959	0.6200	0.6313	0.4520	0.6498	0.7164	0.7436	0.7559
1	2	0.5687	0.6969	0.7416	0.7605	0.7692	0.6226	0.7877	0.8348	0.8529	0.8610
1	3	0.6582	0.7706	0.8074	0.8226	0.8296	0.7122	0.8477	0.8834	0.8969	0.9028
2	1	0.5754	0.7026	0.7468	0.7654	0.7740	0.6226	0.7877	0.8348	0.8529	0.8610
2	2	0.7251	0.8214	0.8516	0.8640	0.8696	0.7674	0.8813	0.9099	0.9206	0.9253
2	3	0.7939	0.8704	0.8934	0.9027	0.9069	0.8319	0.9176	0.9381	0.9456	0.9489
3	1	0.6702	0.7799	0.8156	0.8303	0.8371	0.7122	0.8477	0.8834	0.8969	0.9028
3	2	0.7982	0.8734	0.8959	0.9050	0.9091	0.8319	0.9176	0.9381	0.9456	0.9489
3	3	0.8525	0.9097	0.9263	0.9330	0.9359	0.8813	0.9435	0.9579	0.9631	0.9654

Table 6. Estimates of broad sense heritability (H) for length of 20 seeds (cm) at different stages of inbreeding and combinations of years (y) and repetitions per year (r).

Years (y)	Reps in years (r)	Population 1					Population 2				
		$F_{2:3}$	$F_{3:4}$	$F_{4:5}$	$F_{5:6}$	$F_{6:7}$	$F_{2:3}$	$F_{3:4}$	$F_{4:5}$	$F_{5:6}$	$F_{6:7}$
1	1	0.4651	0.6280	0.6868	0.7118	0.7233	0.4020	0.5946	0.6642	0.6935	0.7069
1	2	0.6046	0.7480	0.7941	0.8128	0.8213	0.5590	0.7344	0.7886	0.8101	0.8198
1	3	0.6718	0.7989	0.8377	0.8532	0.8602	0.6427	0.7969	0.8411	0.8582	0.8658
2	1	0.6349	0.7715	0.8143	0.8316	0.8394	0.5735	0.7457	0.7982	0.8190	0.8283
2	2	0.7536	0.8559	0.8852	0.8968	0.9019	0.7172	0.8469	0.8818	0.8951	0.9010
2	3	0.8037	0.8882	0.9117	0.9208	0.9248	0.7825	0.8870	0.9137	0.9237	0.9281
3	1	0.7229	0.8351	0.8681	0.8811	0.8869	0.6685	0.8148	0.8558	0.8716	0.8786
3	2	0.8210	0.8991	0.9205	0.9287	0.9324	0.7918	0.8924	0.9180	0.9275	0.9317
3	3	0.8600	0.9226	0.9393	0.9458	0.9486	0.8437	0.9217	0.9408	0.9478	0.9509

Table 7. Estimates of broad sense heritability (H) for weight of 20 seeds (g) at different stages of inbreeding and combinations of years (y) and repetitions per year (r).

Years (y)	Reps in years (r)	Population 1					Population 2				
		$F_{2:3}$	$F_{3:4}$	$F_{4:5}$	$F_{5:6}$	$F_{6:7}$	$F_{2:3}$	$F_{3:4}$	$F_{4:5}$	$F_{5:6}$	$F_{6:7}$
1	1	0.3355	0.5216	0.5946	0.6262	0.6409	0.3092	0.4997	0.5756	0.6087	0.6241
1	2	0.4773	0.6635	0.7262	0.7518	0.7635	0.4358	0.6329	0.7007	0.7286	0.7413
1	3	0.5555	0.7296	0.7840	0.8057	0.8154	0.5047	0.6946	0.7554	0.7798	0.7908
2	1	0.5025	0.6856	0.7457	0.7702	0.7812	0.4724	0.6664	0.7307	0.7568	0.7686
2	2	0.6461	0.7977	0.8414	0.8583	0.8659	0.6071	0.7752	0.8240	0.8430	0.8514
2	3	0.7142	0.8437	0.8789	0.8924	0.8983	0.6708	0.8197	0.8606	0.8763	0.8832
3	1	0.6024	0.7659	0.8148	0.8341	0.8426	0.5732	0.7498	0.8027	0.8235	0.8328
3	2	0.7326	0.8554	0.8883	0.9009	0.9064	0.6986	0.8380	0.8754	0.8896	0.8958
3	3	0.7894	0.8901	0.9159	0.9256	0.9298	0.7535	0.8722	0.9026	0.9140	0.9190

Chapter III

Use of Simple Sequence Repeat (SSR) Markers for Mapping Quantitative Trait Loci (QTL) Influencing Early Maturity in *Arachis hypogaea* L.

Franco Villegas Chirinos[2], Thomas G. Isleib*, and Susana R. Milla-Lewis

Formatted for submission to *Crop Science*

[2] F. Villegas Chirinos, T.G. Isleib, and S.R. Milla-Lewis, Dept. of Crop Science, Box 7620, N.C. State Univ., Raleigh, NC 27695-7620. Received: _____. * Corresponding author email: tom_isleib@ncsu.edu

ABSTRACT

In recent years, increased availability of simple sequence repeat (SSR) markers has allowed for the construction of linkage maps and the possibility of marker-assisted selection (MAS) in cultivated peanut (*Arachis hypogaea* L.). Two populations of recombinant inbred lines (RILs), Pop1 (132 RILs) and Pop2 (50 RILs) were constructed from two crosses with one parent in common, Chico-ol_1ol_2, a high-oleic derivative of the early maturing spanish-type PI 565455. The non-common parental lines were PI 313949 for Pop1 and PI 365550 for Pop2, both overo-type late maturing plant introductions. These parental lines were screened for polymorphism with 453 SSR primers previously found to be polymorphic in different cultivated peanut genomes. Of these, 216 (47.7%) markers were polymorphic for at least one of the populations. Linkage maps for Pop1 (86 loci in 22 LG spanning 446.2 cM) and Pop2 (67 loci in 18 LG spanning 284.2 cM) were constructed. A linkage map for the combined data of Pop1 and Pop2 was also created (94 loci in 22 LG spanning 616.4 cM). Phenotypic data for maturity index was collected in replicated field trials for two years and analyzed to identify QTL associated with early maturity. QTL-Maturity I was significant in Pop1 (LOD=4.2 and R^2=0.13) and the combined data (LOD=3.13 and R^2=0.07). QTL-Maturity II was significant only in Pop2 (LOD=2.52 and R^2=0.34), however, increased LOD scores were observed for corresponding regions in Pop1 and the combined data. QTL-Maturity III was only significant in the combined data (LOD=3.69 and R^2=0.10). QTL-Maturity I is the most promising finding for MAS applications in peanut breeding, because it has the highest LOD score, showed consistent location in Pop1 and the combined data, and was not detected only when the smallest population (Pop2, 50 RILs) was used.

INTRODUCTION

Peanut (*Arachis hypogaea* L.) is an economically important legume grown in the U.S.A. Marketed as whole seed for snacks and for the manufacture of products such as peanut butter and candy (USDA FAS, 2011), it is also crushed for oil when quality standards for direct consumption are not met. In the U.S.A., peanuts are often grown at high latitudes, regions such as northern North Carolina, southern Virginia, and west Texas that are often affected by end-of-season frosts or night temperatures cold enough to retard the maturation process and induce incomplete seed development. All the mentioned regions would benefit from deployment of early-maturing peanut cultivars that could avoid the yield and grade penalties derived from short growing seasons.

It is difficult to assess the maturity of peanuts because pods grow below the soil surface and because the species' indeterminate flowering creates a wide distribution of maturity levels among the pods on a single plant. Furthermore, maturity level is a quantitative trait with reported low heritability (Chiow and Wynne, 1983; Ali *et al.*, 1999), and influenced by many genes and the environment (Pattee *et al.*, 1974, Rowland *et al.*, 2006). The currently preferred assessment method, the hull-scrape method (William and Drexler, 1981), involves the laborious removal of the peanut exocarp, followed by classification of pods into maturity classes based on mesocarp color, a process whose results can vary with different observers. Because of this difficulty in assessment, alternative methods with higher efficiency and less subjectivity would be desirable for selection of early maturity in peanuts. Marker-assisted selection (MAS) is an indirect selection method, where molecular markers associated with a trait, are used to identify desirable genotypes. A linkage map with marker density high

enough to allow identification of genomic regions associated with early maturity is necessary for successful MAS. Other methods for which no QTL mapping is required such as genome-wide selection are not available for peanuts due to a paucity of informative markers for this crop (Bernardo and Yu, 2007). Until recently, molecular breeding techniques in cultivated peanut were hindered due to the lack of DNA markers with sufficient variability among cultivars and landraces of *A. hypogaea*. This homogeneity probably originated from the recent formation of the allotetraploid peanut genome (AABB, 2n=4x=40), where one or a few sexually isolated hybrid plants resulted from the fusion of two related genomes. Because of this paucity in molecular variability, the first attempts to produce linkage maps in peanuts were made using wild *Arachis* species or crosses of cultivated peanuts with synthetic tetraploid interspecific amphiploids. Using these kinds of mapping populations, much higher levels of allelic diversity and linkage disequilibrium were found compared to crosses within *A. hypogaea*. Several studies have produced linkage maps using F_2 (Halward *et al.*, 1993; Moretzshon *et al.*, 2005; Leal-Bertoli *et al.*, 2009), and backcross (Burrow *et al.*, 2001; Fonceka *et al.*, 2009) mapping populations with diploid wild species of the genus *Arachis* (Halward *et al.*, 1993; Fonceka *et al.*, 2009; Moretzshon *et al.*, 2005) or a synthetic amphiploid (Burow *et al.*, 2001; Fonceka *et al.*, 2009), as parents. Mapping of leaf spot resistance using a linkage map based on an F_2 population from a cross of two diploid species was performed by Leal-Bertoli *et al*. (2009). They found several QTL that could be useful for introgression purposes.

In recent years, several hundred new SSR markers have been developed for peanuts (Varshney *et al.*, 2007; Moretzsohn *et al.*, 2004; Ferguson *et al.*, 2004; He *et al*., 2003;

Hopkins *et al.*, 1999). SSRs are highly informative, multiallelic, codominant markers that are more polymorphic than AFLPs or RAPDs in *A. hypogaea* (He *et al.*, 2003). Furthermore, SSR markers are based on simple and highly reproducible PCR reactions and show high transferability among related species (Appleby *et al.*, 2009). The recent availability of a critical mass of informative SSR markers in *A. hypogaea* allowed for construction of linkage maps with sufficient marker density to consider QTL mapping and subsequent MAS. In the last three years, four projects (Hong *et al.*, 2008; Varshney *et al.*, 2009; Hong *et al.*, 2010; Khedikar *et al.*, 2010) constructed linkage maps for *A. hypogaea*, all of them using SSR markers. Varshney *et al.* (2009) mapped traits related to drought resistance including transpiration (T), transpiration efficiency (TE), specific leaf area (SLA), and SPAD chlorophyll meter reading (SCMR), identifying four QTL for T, four for TE, 14 for SLA, and six for SCMR. However, the phenotypic variation explained by these QTL was low, ranging from 3.5 to 14.1%. The authors concluded that none of the identified QTL had potential for use in MAS breeding. Khedikar *et al.* (2010) mapped late leaf spot (LLS) and rust resistance using composite interval analysis (Zeng, 1994). Eleven QTL associated with LLS explained 1.7-6.5% of the phenotypic variation, and 12 QTL associated with rust were found to explain 1.7-55.2% of the phenotypic variation. One QTL, QTL(rust)01, explained 55.2% of the variability in rust resistance level, making it a promising discovery for introgression purposes using marker-assisted backcrossing.

Recent progress in the development of genomic data necessary for molecular breeding for *A. hypogaea* is evident, making the future of MAS for this crop promising. In the last three years, four projects worked to construct easily comparable SSR-based linkage maps for

cultivated peanuts with variable marker density and resolution (Hong *et al.*, 2008; Varshney *et al.*, 2009; Hong *et al.*, 2010; Khedikar *et al.*, 2010). Moreover, two of those projects already used their maps for QTL mapping of important drought and disease resistance traits. However, low marker density is still a problem for mapping purposes. The main objectives of this study were 1) to increase the amount of genotypic data involving linkage map construction with SSRs for *A. hypogaea*, putting the peanut community a step closer to a comprehensive, high resolution, easily comparable linkage map for such an important crop, and 2) to locate regions in the peanut genome associated with early maturity for the development of alternative MAS breeding strategies for this important trait with complex heredity patterns.

MATERIALS AND METHODS

Population Development. Two mapping populations of recombinant inbred lines (RILs) were developed using parents differing widely in maturation times and other agronomic traits. Population 1 (Pop1) consists of 132 F_6-derived families produced from the cross of plant introduction PI 313949 with a high-oleic backcross derivative of PI 565455 (hereafter designated "Chico-*ol*$_1$*ol*$_2$"). PI 565455 is a spanish-type (*A. hypogaea* subsp. *fastigiata* Waldron var. *vulgaris* Harz) germplasm line selected from PI 268661, presented to the National Plant Germplasm System by the Mount Makulu Research Station, Chilanga, Rhodesia (now Zambia) in 1960 (Bailey and Hammons, 1975). The line was sent to Rhodesia by V.S. Pustovoit from the All-Russia Research Institute of Oil Crops (VNIIMK) in Krasnodar in the former Union of Soviet Socialist Republics (now the Russian Federation) where it had the designation "APAXUC [*Arachis*] 370". PI 313949 is a Bolivian overo-type peanut (*A. hypogaea* subsp. *hypogaea* var. *hypogaea*) with normal-oleic seed oil, late maturity (it is ready for harvest around 130 days after planting, DAP), large pods, and variegated red-and-white seeds (USDA GRIN, 2011a, USDA GRIN, 2011b). Chico-*ol*$_1$*ol*$_2$ is a high-oleic, very early maturing (usually 100 DAP) spanish type line with small pods and seeds (USDA GRIN, 2011c). Population 2 (Pop2) consists of 50 F_6-derived families derived from the cross of PI 365550 with the same parent as Pop1, Chico-*ol*$_1$*ol*$_2$. PI 365550 is another Bolivian overo-type line, with normal-oleic seed oil, very late maturity, and very large pods and seeds (USDA GRIN, 2011d; USDA GRIN, 2011e).

All crosses were made at the NCSU greenhouse facility in the summer of 2000. The F_1 plants were grown at the Peanut Belt Research Station (PBRS) near Lewiston, NC in 2001, the F_2 populations in 2002. Individual F_2 plants were harvested. The F_2-derived families were advanced by the modified pedigree method (single seed descent). In 2007, two random F_6 plants were harvested from each F_2-derived family forming the two populations used in this study. The 182 F_6-derived families were planted with the three parents and 13 additional checks in a 14x14 simple square lattice design at PBRS in 2008 and 2009. Plots comprised two rows 3.7 m in length, spaced 91 cm apart with 25 cm spacing between seeds at planting. The plots were irrigated and provided with a full program of chemical control of weeds and diseases. The plots were dug 125 DAP the first year, and 130 DAP the second year. A single pod was hand-picked from each plant in each plot of the first replicate, and then the plots were harvested using a stationary thresher. After recording yields, a single bag of approximately 1 kg of pods from each individual bulk-harvested plot was stored for phenotypic evaluation at NCSU.

Phenotypic Evaluation and Data Analysis. Maturity was assessed using the hull-scrape method described by Williams and Drexler (1981). Briefly, the exocarp of a sample of 100 peanut pods were removed using a Task Force® 2000 PSI electric high-pressure washer (Rexon Industrial Corp, Taiwan, ROC), with 2000 maximum PSI, 1.6 GPM maximum flow rate, and a turbo spray nozzle with an orifice size of 1.5 mm. The samples were treated for approximately 20 seconds with the maximum pressure to ensure accurate measurements. Subsequently, the ratio of the number of pods in the black and brown mesocarp classes (mature fruits) over the total number of blasted pods was used as an

estimator of maturity called "maturity index." All phenotypic measurements were carried out for two years, 2008 and 2009. Phenotypic data for the two years were analyzed using the general linear models procedure (PROC GLM) of SAS statistical software Version 9.2 (SAS Institute, Cary, NC). All effects were considered to be fixed, and F_6-derived family means were computed.

SSR Genotyping. Four to ten of the seeds representing separate $F_{6:7}$ plants from the 2008 test were grown under greenhouse conditions and sampled to provide DNA for marker analysis. Small sections of approximately a third of the youngest leaf of each plant were collected, at the six-leaf stage of growth, to form a pooled sample for each F_6-derived family. This pooled sample was used in order to have a high probability of detecting all existing alleles in a family for any molecular markers still heterozygous in the F_6 plants and for the genes controlling the traits studied. Total genomic DNA was extracted using a CTAB extraction protocol as described by Stein *et al.* (2001). The 196 DNA samples were stored at -70°C. The three parents used to develop the populations were sent to the Univ. of Georgia-Athens to screen a set of 453 SSR markers for polymorphism. This set was selected from a larger group of 2694 markers that were previously identified as polymorphic using cultivated tetraploid peanut genotypes (S.J. Knapp, personal commun.). The selected 453 markers were polymorphic between at least two different cultivated peanut genotypes. Subsequently, these SSR markers were screened with the three parental genotypes used in our study.

PCR reactions, size fractionation, and allele sizing for the SSR markers were performed as described by Varshney *et al.* (2009). PCR reactions for the SSR genotyping was performed in 10 µl reaction volumes containing 4 pmoles of primers (with the forward

primer labeled with one of three fluorescence dyes 6-FAM, 5-HEX or TAMRA, and the reverse primer without label), 10 mM $MgCl_2$, 3.3 mM dNTPs, 0.5 U of Taq DNA polymerase (New England Biolabs, Ipswich, MA), and 1X PCR buffer (New England Biolabs, Ipswich, MA). A touch-down PCR amplification protocol with 1 min of initial denaturation, followed by six cycles of 94ºC for 30 s, 64ºC for 25 s and 72ºC for 50 s, with 1ºC decrease in annealing temperature per cycle, then 33 cycles of 94ºC for 20 s, with a constant annealing temperature of 57ºC for 20 s and 72ºC for 50 s, followed by a final extension at 72 ºC for 15 min was used in an Eppendorf Mastercycler ep384 (Eppendorf, Hauppauge, NY). PCR products were size-fractioned using capillary electrophoresis on an ABI 3730 automatic DNA sequencer (Applied Biosystems, Carlsbad, CA). Allele sizing of the electrophoretic data thus obtained was performed using Genemarker v. 1.85 (Softgenetics, State College, PA).

Linkage map construction and QTL mapping. The segregation pattern for each individual marker was subjected to a chi-square (χ^2) test to detect distortion from the expected 1:1 segregation ratio. Linkage analysis was performed using JoinMap v.4.0 (Van Oojen and Voorips, 2001). Genotypic and phenotypic data were analyzed in order to identify QTL associated with the traits of interest by using the composite interval mapping method proposed by Zeng (1994) in WinQTL Cartographer, version 2.5. (Wang, 2007) as described by Varshney *et al.* (2009).

RESULTS AND DISCUSSION

SSR Marker Polymorphism. Of the original 453 SSR markers, 177 (37.1%) were polymorphic between Chico-*ol*$_1$*ol*$_2$ and both PI 313949 and PI 36550. Twenty four (5.3%)

were polymorphic between Chico-*ol*$_1$*ol*$_2$ and only PI 313949; and 15 (3.3%) between Chico-*ol*$_1$*ol*$_2$ and only PI 365550. That is 216 (47.7%) markers were polymorphic for at least one of the populations used in our study. This is a very high proportion of polymorphic markers for cultivated tetraploid peanut genomes considering that scarce genetic variability at the DNA marker level has been repeatedly reported within *A. hypogaea* species in the past (Kochert *et al.*, 1991; Hilu and Stalker, 1995; Subramanian *et al.*, 2000; Gimenes *et al.*, 2002; Milla *et al.*, 2005). A possible explanation for this high rate of polymorphism is that the markers used were selected from a larger set of 2694 SSRs which already had been found to be polymorphic between at least two cultivated peanut genomes in a previous study (Ma *et al.*, 2006). Despite having been screened using genotypes different from the ones used here, this pre-screening seems to have enriched for SSR markers that tend to be polymorphic among cultivated peanut genomes. The overo-type parents and the spanish-type parent were collected in very different environments and showed many morphological and agronomic differences. Therefore, genetic variability within the hybrid populations is expected to be higher compared with crosses of more closely related parents. Furthermore, during the parental screening, 330 (72.8%) markers did not show any difference in allele size between PI 313949 and PI 365550, suggesting a high degree of similarity at the DNA level between these genotypes and a comparatively high degree of dissimilarity between the overo-type group and Chico-*ol*$_1$*ol*$_2$.

Linkage map construction. Linkage analysis was performed taking into consideration the family structure of the populations used. The 132 F_6-derived families in Pop1 can be traced to 66 F_2-derived families, so that each F_2-derived family is represented by two F_6-

derived families. Therefore, we randomly chose only one of the two F_6-derived families per F_2-derived family, obtaining 66 lines total for linkage map construction. We repeated this process 20 times and compared the results. Each time the selection between the two F_6 lines from the same F_2 was random and independent from previous selections. We did not test complementary sets. Linkage groups were very consistent across all 20 maps, with almost always the same markers being included in a particular linkage group (LG) and marker positions being highly conserved in most cases. Considering that the objective of this study was to find QTL associated with early maturity, and that having more marker information increases the power for detecting QTL (Z.B. Zeng, personal commun.), we decided to include all 132 lines and produce a single linkage map per population. This map had the same number of LGs but included 13 more markers than the average map constructed with 66 F_6-derived lines. However, linkage groups were highly conserved in number and position with the average map (data not shown). The same reasoning was applied to Pop2 with 50 RILs. However, the smaller number of lines in this population increased the variability of the estimated recombination frequencies. Therefore, more variability in marker position was expected when comparing 20 maps from 25 randomly chosen F_6-derived families.

Linkage Map for Pop1 (PI 313949 x Chico-ol$_1$ol$_2$). The original number of polymorphic SSR markers for Pop1 was 201. Out of these, 90 were not scored due to poor amplification, fluorophore degradation, or calibration issues with the sequencer. Therefore, 111 SSR markers, amplifying 116 different loci, were used to construct a linkage map that included 22 LG, and spanned 446.2 cM (Fig. 1). An independence LOD value of 3 was used as threshold to define LG. Thirty loci remained ungrouped at this LOD value. The LG sizes ranged from

3.36 to 72.99 cM, and the number of loci per LG ranged from 2 to 8. The average distance between loci was 5.19 cM. Forty four loci, 51.2% of the 86 included in the linkage map, showed segregation distortion from the expected 1:1 ratio (for co-dominant markers in RILs). The number of loci showing segregation distortion per LG ranged from zero to five, with only LG6 not showing a distorted locus. Thirty-four (77%) of the distorted loci were skewed toward the PI 313949 parental allele. The most extreme case of segregation distortion was found in LG12 where all five loci favored the PI 313949 parental allele. This LG had an average ratio of 171 alleles from Chico-ol_1ol_2 to 411 alleles from PI 313949. That means there is a 70.6% probability of randomly choosing an overo allele when screening loci of the F_6-derived families (instead of the theoretical 50% for non-distorted loci). These results suggest that natural selection operated in the population during development of $F_{2:6}$ families and favored PI 313949 alleles. Potential sources of natural selection could be a particular disease, such as web blotch (caused by *Phoma arachidicola* Marasas, Pauer, & Boerema), for which spanish type (fastigiate) genotypes are extremely susceptible (Phipps, 1985). Another disease, observed at very high levels during the growing seasons of 2008 and 2009, was tomato spotted wilt virus (TSWV).

Linkage Map for Pop2 (PI 365550 x Chico-ol_1ol_2). From the original 192 SSR markers for Pop2, only 107 were scored due to problems similar to the ones found in Pop1. These markers amplified 113 different loci that were used to construct a linkage map with 18 LGs and spanning 284.2 cM (Fig. 2). Again, an independence LOD value of 3.0 was used as threshold for forming LGs. Forty-six loci remained ungrouped at this LOD score. The LG sizes varied in size from <0.001 to 51 cM, and each LG had from two to seven loci. The

average distance between loci was 4.24cM. Thirty-five (52.2%) of the 67 loci forming the LGs presented segregation distortion. From zero to seven loci per LG showed segregation distortion. LGs 2, 7, 10, and 13 did not show a locus with segregation distortion. Twenty seven of the 35 distorted markers (77.1%) favored the allele from PI 365550. LG1 presented the most severe case of segregation distortion with all seven loci skewed towards the PI 365550 allele. This LG had an average ratio of 72 alleles from Chico-ol_1ol_2 to 249 alleles from PI 313949 (the probability of randomly recovering the overo allele was 77.6%). Again the alleles from the overo-type parent, PI 365550, were favored over those of the spanish-type parent, Chico-ol_1ol_2. The natural selection pressure described for Pop1 may have caused the observed segregation distortion in Pop2 also.

Linkage Map for the combined data of Pop1 and Pop2. The populations used in this study had one parent in common. The female parents, PI 313949 and PI 365550, were very similar for agronomic traits such as fruit and seed dimensions, growth habit, and number of seeds per pod. However, the contrast between maturity indexes for PI 313949 and PI 365550 was reported significant (P=0.0077) in a parallel study (Villegas, 2011). Nevertheless, PI 313949 and PI 365550 belong to the overo peanut type, and were collected from Bolivia (USDA GRIN, 2011a; USDA GRIN, 2011d). Furthermore, they were highly conserved at the molecular marker level with the same allele sizes for 72.8% of the markers screened. However, Pop1 and Pop2 had extremely different population sizes (132 vs. 50 RILs). Given that there is increased variability in recombination frequencies estimated from smaller population sizes, making them less reliable, data from all 182 lines were combined. Loci with different allele sizes in the overo-type parents were considered missing data for Pop2,

resulting in the loss of 37 x 50 data points. In this way, a linkage map with 22 LGs, and spanning a distance of 616.4 cM was created (Fig. 3). A threshold of independence LOD value of 3 or higher was used. Twenty-two loci remained ungrouped at this LOD score. LG sizes were within 3.2 and 74.4 cM, with two to eight loci per LG. The average distance between loci was 6.6 cM. Sixty-one (64.9%) of 94 loci included in the map exhibited segregation distortion. The number of distorted markers per LG ranged from one to six. Thirty-six (59%) of the 61 distorted loci favored the overo-type allele over the Chico-ol_1ol_2 allele. The LG with the highest number of distorted loci was LG4, with six distorted loci of a total of eight. In this LG, five of the distorted loci were skewed toward the Chico-ol_1ol_2 allele, but only one locus was extremely distorted toward the overo-type allele (36 Chico-ol_1ol_2 allele : 127 overo-type allele, that is 77.9% fixation of the overo allele). The average ratio of Chico-ol_1ol_2 allele : overo-type allele for LG5 was only 695 : 596 favoring the Chico-ol_1ol_2 alleles.

Phenotypic data: Maturity Index. The mean maturity indices for RILs and parental genotypes were obtained from data from two years with two repetitions per year. All genotypes were grown at the PBRS in 2008 and 2009. PI 313949, PI 365550, and Chico-ol_1ol_2 had maturity indexes of 42.9%, 17.5%, and 96.2% respectively. The contrast between PI 313949 and Chico-ol_1ol_2, was highly significant ($P<0.0001$) as was that between PI 365550 and Chico-ol_1ol_2 ($P<0.0001$). The contrast between the overo-type parents was significant ($P=0.0077$). Variability among RILs in each population was highly significant ($P<0.0001$). Maturity indices ranged from 0.38 to 0.90 in Pop1 and 0.43 to 0.91 in Pop2. These results are evidence of high variability in maturity indexes in both populations

evaluated in this study. The median maturity index was 0.68 in Pop1, 0.63 in Pop2, and 0.67 in Pop1 + Pop2 (Fig. 4).

QTL analysis. Each of the previously described linkage maps, together with the corresponding maturity level information for the lines involved in map construction were used to identify regions of the peanut genome associated with this trait. Using only the data obtained for Pop1 (132 RILs), one QTL was detected (Fig. 5). This QTL, designated "QTL-Maturity I", had a peak value located between markers GM34 and GM2689, a LOD value of 4.17, and an R^2 maximum value of around 0.13. When the combined data for Pop 1 + Pop 2 were used, QTL-Maturity I was again identified between markers GM34 and GM2689 (Fig. 6). This time the QTL had a maximum LOD value of 3.13 and a maximum R^2 value of 0.07. When data only for Pop2 was used, the equivalent regions associated with QTL-Maturity I (including markers GM34 and GM2689) did not show any increase in LOD score. Moreover, when data for Pop 2 was used, a QTL, designated "QTL-Maturity II", with a LOD value of 2.52 and an R^2 value of 0.34 (Fig. 7) was detected between markers GM2165 and GM2032. This QTL was barely significant in Pop2 and was significant neither in Pop1 nor the combined data. However, despite not reaching their respective threshold values, increased LOD values for regions close to the corresponding LGs for Pop1 (peak at LOD=2.34 and R^2=0.09) and Pop1+Pop2 (peak at LOD=2.14 and R^2=0.07) were observed (Figs. 8 and 9). This suggests the presence of QTL-Maturity II is real and not specific for Pop2. However, this QTL might be difficult to detect because it has a true small effect, as suggested by the R^2 from the more informative Pop1 and the combined population analysis. The R^2 obtained in Pop2 seem to be inflated due to its small population size. Further testing

with larger population sizes is necessary to confirm this QTL. A third QTL, "QTL-Maturity III" (Fig.10), with LOD 3.65 and R^2 value of 0.07 was detected when using the combined data. However, this QTL was not detected using either population alone, suggesting that it could be an artifact produced from the fusion of Pop1 and Pop2. We considered as missing data for Pop2 all loci with different allele sizes for the overo-type parents. This procedure seems to have reduced the LOD value of the detected QTL. Redoing the analysis with the combined populations considering all the overo-type alleles as common regardless of allele size would improve the power of the analysis, producing higher LOD scores for the detected QTL (data not shown). However, the theoretical implications of such analysis would be very difficult to manage.

QTL-Maturity I and its associated markers, GM34 and GM2689, are the most promising finding in terms of future MAS applicability, because this QTL showed the highest LOD score (significance level), had an R^2 value of 0.14 in Pop1, was consistently located between the mentioned markers in data from Pop1 and the combined population, and was not detected only when data from the smallest population (Pop2, 50 RILs) was used. We must bear in mind that smaller population sizes produces more unreliable results for QTL detection (the estimated statistics have more standard deviation). QTL-Maturity II was only significant in the smallest population (Pop2), however, the corresponding regions in Pop1 and the combined data had LOD scores that were elevated but did not reach the threshold value. QTL-Maturity II, seem to have a small true effect (R^2 for Pop1 is 0.09 and for Pop1+Pop2 is 0.07) that could have been falsely inflated in Pop2 (R^2=0.34) due to increased variability of the estimates. It is our opinion that QTL-Maturity-II has a small chance for

successful MAS application. However, fine mapping with more SSR markers within the region associated with this QTL may enhance its real R^2 value increasing its chances for applicability. A MAS program for early maturity could not only increase the response to selection but also allow for selection in off-season nurseries enabling the breeder to better allocate resources after discarding lines with little promise. Perhaps, a "mixed" breeding program, with phenotypic selection on target locations and MAS (with very low selection intensity) in off-site locations would be the best starting point for testing the feasibility of MAS for early maturity in peanuts.

Acknowledgements;

The authors thank the Center for Applied Genetic Technologies of the University of Georgia Athens, Dr. Steve Knapp, and the Eastern Regional Small Grains Genotyping Lab (ERSGGL) directed by Dr. Gina Brown (USDA-ARS); for their collaboration in the data acquisition steps of this study.

REFERENCES

Ali, N., J.C. Wynne, and J.P. Murphy. 1999. Estimation of genetic effects and heritability for early maturity and agronomic traits in peanuts (*Arachis hypogaea* L.). Pak. J. Bot., 31(2): 323-335.

Appleby, N., D. Edwards, and J. Batley. 2009. New technologies for ultra-high throughput genotyping in plants. p. 19-39. *In* D.J. Somers, P. Langridge, and J.P. Gustafson (eds.), Methods in Molecular Biology, Vol 513: Plant Genomics methods and protocols. Humana Press. New York, NY.

Bailey, W.K., and R.O. Hammons. 1975. Registration of Chico peanut germplasm. Crop Sci. 15: 105.

Bernardo, R. and J. Yu. 2007. Prospects for Genome-wide selection for quantitative traits in maize. Crop Science 47: 1082-1090.

Burow, M.D., C.E. Simpson, J.L. Starr, and A.H. Patterson. 2001. Transmission genetics of chromatin from a synthetic amphidiploid to cultivated peanut (*Arachis hypogaea* L.): broadening the gene pool of a monophyletic polyploidy species. Genetics 159:823-837.

Chiow, H.Y. and J.C. Wynne. 1983. Heritabilities and genetic correlations for yield and quality traits of advanced generations in a cross of peanut. Peanut Sci. 10:13-17.

Ferguson, M.E., M.D. Burow, S.R. Schulze, P.J. Bramel, A.H. Paterson, S. Kresovich, and S. Mitchell. 2004. Microsatellite identification and characterization in peanut (*A. hypogaea* L.). Theor. Appl. Genet. 108: 1064–1070.

Fonceka, D., T. Hodo-Abalo, R. Rivallan, I. Faye, M. N. Sall, O. Ndoye, A. P. Favero, D. J. Bertiolis, J.C. Glaszmann, B. Courtois, and J.F. Rami. 2009. Genetic mapping of wild

introgressions into cultivated peanut: a way toward enlarging the genetic basis of a recent allotetraploid. BMC Plant Biology 9:103.

Gimenes, M.A., C.R. Lopes, and J.F.M. Valls. 2002. Genetic relationships among *Arachis* species based on AFLP. Genet. Mol. Biol., 25:349-353.

Halward, T. M., H.T. Stalker, and G. Kochert. 1993. Development of an RFLP linkage map in diploid peanut species. Theor. Appl. Genet. 87: 379-384.

He, G., R. Meng, M. Newman, G. Gao, R.N. Pittman, and C.S. Prakash. 2003. Microsatellites as DNA markers in cultivated peanut (Arachis hypogaea L). BMC Plant Biology 3:3.

Hilu, K.W. and H.T. Stalker. 1995. *G*enetic relationships between peanut and wild species of *Arachis* sect. *Arachis* (*Fabaceae*): evidence from RAPDs. Plant. Syst. Evol., 198: 167-178.

Hong, Y., X. Liang, X. Chen, H. Liu, G. Zhou, S. Li, and S. Wen. 2008. Construction of Genetic Linkage Map Based on SSR Markers in Peanut (*Arachis hypogaea* L.). Agricultural Sciences in China, 7: 915-921.

Hong, Y., X. Chen, X. Liang, H. Liu, G. Zhou, S. Li, S. Wen, C.C. Holbrook, and B. Guo. 2010. A SSR-based composite genetic linkage map for the cultivated peanut (*Arachis hypogaea* L.) genome. BMC Plant Biology, 10: 17.

Hopkins, M.S., A.M. Casa, T. Wang, S.E. Mitchell, R.E. Dean, G.D. Kochert, and S. Kresovich. 1999. Discovery and characterization of polymorphic simple sequence repeats (SSRs) in peanut. Crop Sci. 39: 1243-1247.

Khedikar, Y.P., M.V.C. Gowda, C. Sarvamangala, K.V. Patgar, H.D. Upadhyaya, and R.K. Varshney. 2010. A QTL study on late leaf spot and rust revealed one major QTL for molecular breeding for rust resistance in groundnut (*Arachis hypogaea* L.). Theor. Appl. Genet. 121: 971-984.

Kochert, G., T. Halward, W.D. Branch, and C.E. Simpson. 1991. RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. Theor. Appl. Genet. 81: 565-570.

Leal-Bertioli, S.C.M., A.C.V.F. Jose, D.M.T. Alves-Freitas, M.C. Moretzsohn, P.M. Guimaraes, S. Nielen, B.S. Vidigal, R.W. Pereira, J. Pike, A.P. Favero, M. Parniske, R.K. Varshney, and D.J. Bertioli. 2009. Identification of candidate genome regions controlling disease resistance in *Arachis.* BMC Plant Biology, 9:112.

Ma, W., Y. Li, B. Guo, A.K. Culbreath, S.R. Milla-Lewis, S.P. Tallury, C.C. Holbrook, T.G. Isleib, H.T. Stalker, and S.J. Knapp. 2006. Simple sequence repeat polymorphisms in peanut. *In*: Proceedings of the Plant and Animal Genome XV Conference, January 13-17, 2007, San Diego, CA. pp. 13.

Milla, S.R., T.G. Isleib, and H.T. Stalker. 2005. Taxonomic relationships among *Arachis* sect. *Arachis* species as revealed by AFLP markers. Genome. 48: 1-11.

Moretzsohn, M.C., M.S. Hopkins, S.E. Mitchell, S. Kresovich, J.F.M. Valls, and M.E. Ferreira. 2004. Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. BMC Plant Biology, 4: 11.

Moretzsohn, M.C., L. Leoi, K. Proite, P.M. Guimaraes, S.C.M. Leal-Bertioli, M.A. Gimenes, W.S. Martins, J.F.M. Valls, D. Grattapaglia, and D.J. Bertioli. 2005. A

microsatellite-based, gene-rich linkage map for the AA genome of *Arachis* (*Fabaceae*). Theor. Appl. Genet. 111: 1060-1071.

Pattee, H.E., E.B. Johns, J.A. Singleton, and T.H. Sanders. 1974. Composition changes of peanut fruit parts during maturation. Peanut Sci. 1: 57–62.

Phipps, P.M. 1985. Web blotch of peanut in Virginia. Plant Disease 69: 1097-1099.

Rowland, D.L., R.B. Sorensen, C.L. Butts, and W.H. Faircloth. 2006. Determination of maturity and degree day indices and their success in predicting peanut maturity. Peanut Sci. 33: 125–136.

Subramanian, V., S. Gurtu, R.C. Nageswara Rao, and S.N. Nigam. 2000. Identification of DNA polymorphism in cultivated groundnut using random amplified polymorphic DNA (RAPD) assay. Genome, 43: 656-660.

Stein, N., G. Herren, and B. Keller. 2001. A new DNA extraction method for high-throughput marker analysis in a large genome species such as *Triticum aestivum*. Plant Breeding 120: 354-356.

USDA-FAS Foreign Agricultural Service. 2011. Table 13: Peanut Area, Yield, and Production. Available at http://www.fas.usda.gov/psdonline/psdReport.aspx?hidReportRetrievalName=Table+13+Peanut+Area%2c+Yield%2c+and+Production&hidReportRetrievalID=918&hidReportRetrievalTemplateID=1.

USDA-ARS National Genetic Resources Program. 2011a. Germplasm Resources

Information Network (GRIN) database. Arachis hypogaea L. FABACEAE 'Goober'. Available at http://www.ars-grin.gov/cgi-bin/npgs/acc/display.pl?1234443. National Germplasm Resources Laboratory, Beltsville, MD.

USDA-ARS National Genetic Resources Program. 2011b. Germplasm Resources Information Network (GRIN) database. Observations for accession PI 313949. Available at http://www.ars-grin.gov/cgi-bin/npgs/acc/obs.pl?1234443. National Germplasm Resources Laboratory, Beltsville, MD.

USDA-ARS National Genetic Resources Program. 2011c. Germplasm Resources Information Network (GRIN) database. Observations for accession PI 565455. Available at http://www.ars-grin.gov/cgi-bin/npgs/acc/obs.pl?1460426. National Germplasm Resources Laboratory, Beltsville, MD.

USDA-ARS National Genetic Resources Program. 2011d. Germplasm Resources Information Network (GRIN) database. Arachis hypogaea L. FABACEAE 'Goober'. Available at http://www.ars-grin.gov/cgi-bin/npgs/acc/display.pl?1273705. National Germplasm Resources Laboratory, Beltsville, MD.

USDA-ARS National Genetic Resources Program. 2011e. Germplasm Resources Information Network (GRIN) database. Observations for accession PI 365550. Available at http://www.ars-grin.gov/cgi-bin/npgs/acc/obs.pl?1273705. National Germplasm Resources Laboratory, Beltsville, MD.

Van Oojen J.W. and R.E. Voorips. 2001. *JoinMap*® Version 3.0, Software for the calculation of genetic linkage maps. Plant Research International, Wageningen. The Netherlands.

Varshney, R.K., D.A. Hoisington, H.D. Upadhyaya, P.M. Gaur, S.N. Nigam, K. Saxena, V. Vadez, N. K. Sethy, S. Bhatia, R. Aruna, M.V. Channabyre Gowda, and N.K. Singh. 2007. Molecular genetics and breeding of grain legume crops for the semi arid tropics, p. 207-242 *In* R.K. Varshney, and R. Tuberosa (eds) Genomics-Assisted Crop Improvement, *vol 2* : Genomics Applications in Crops , Springer, Dordrecht, The Netherlands.

Varshney, R.K., D.J. Bertioli, M.C. Moretzsohn, V. Vadez, L. Krishnamurthy, R. Aruna, S.N. Nigam, B.J. Moss, K. Seetha, K. Ravi, G. He, S.J. Knapp, and D.A. Hoisington. 2009. The first SSR-based genetic linkage map for cultivated groundnut (*Arachis hypogaea* L.). Theor. Appl. Genet. 118: 729-739.

Villegas Chirinos, F. 2011. Breeding for Early Maturity in Peanuts (*Arachis hypogaea* L.) using Traditional Methods and Marker Assisted Selection (MAS). M.S. Thesis. North Carolina State University, Raleigh, NC.

Wang, S., C.J. Basten, and Z.B. Zeng. 2007. Windows QTL Cartographer 2.5. http://statgen.ncsu.edu/qtlcart/WQTLCart.htm. Cited 2007.

Williams, E. J. and J. S. Drexler. 1981. A non-destructive method for determining peanut pod maturity. Peanut Sci. 8: 134-141.

Zeng, Z.B. 1994. Precision mapping of quantitative trait loci. Genetics 136:1457-1468.

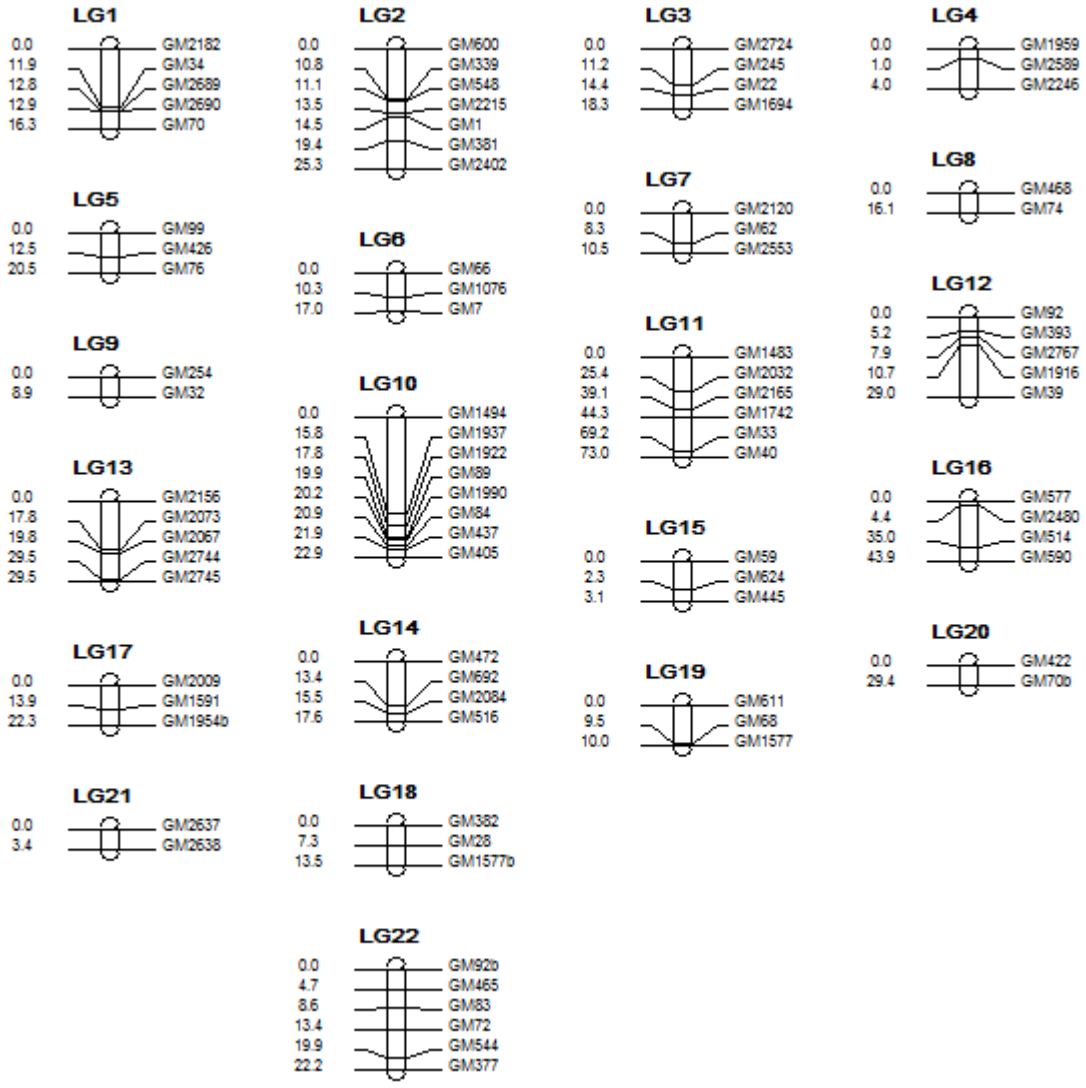

Figure 1. Linkage map created with Pop1 (132 RILs) including 86 loci.

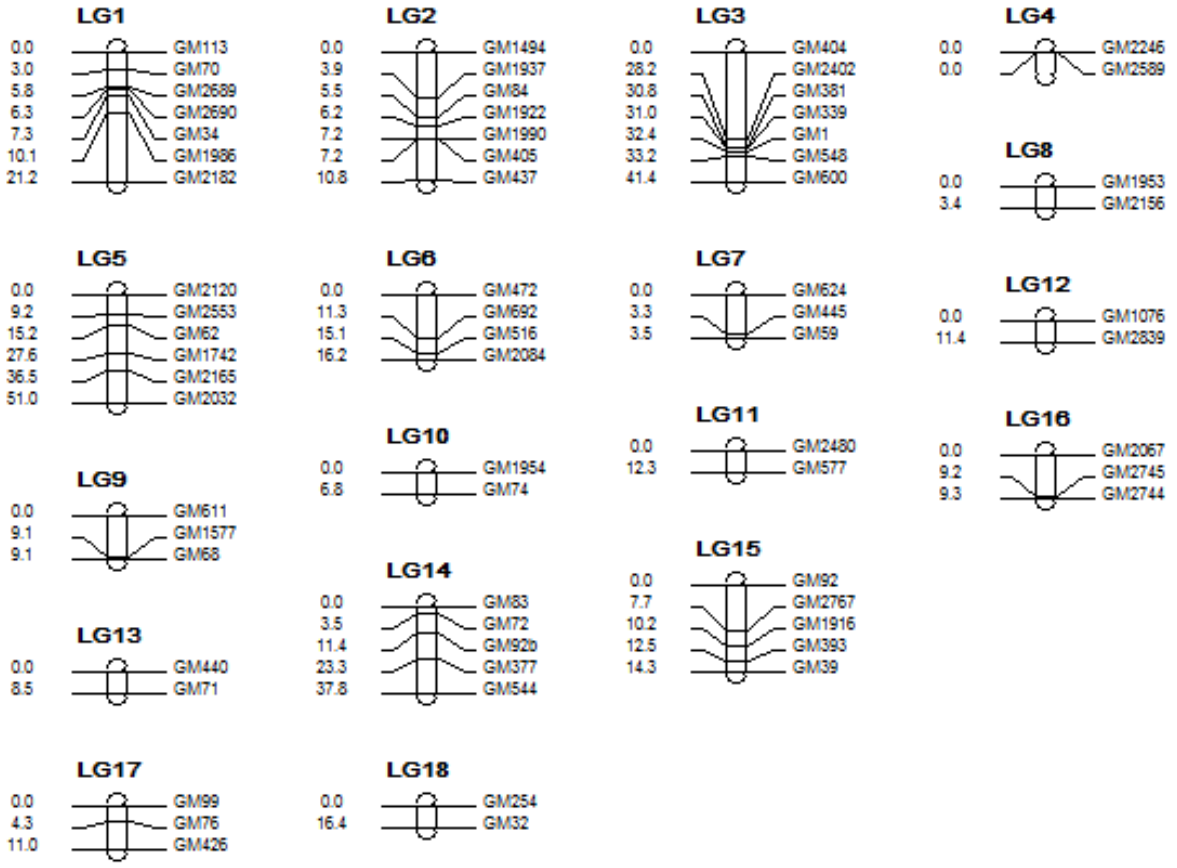

Figure 2. Linkage map created with Pop2 (50 RILs) including 67 loci.

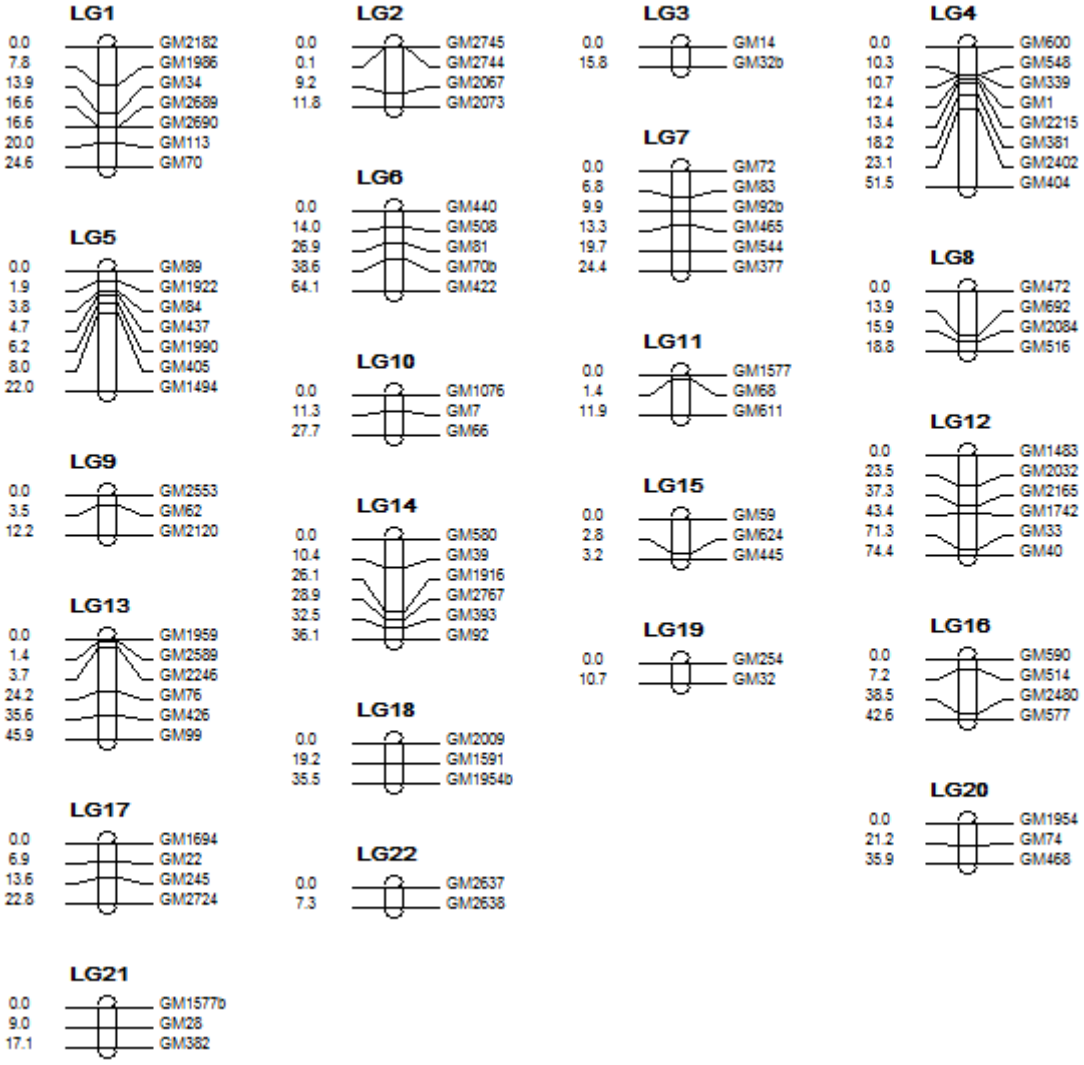

Figure 3. Linkage map created with Pop1 + Pop2 (182 RILs) including 94 loci.

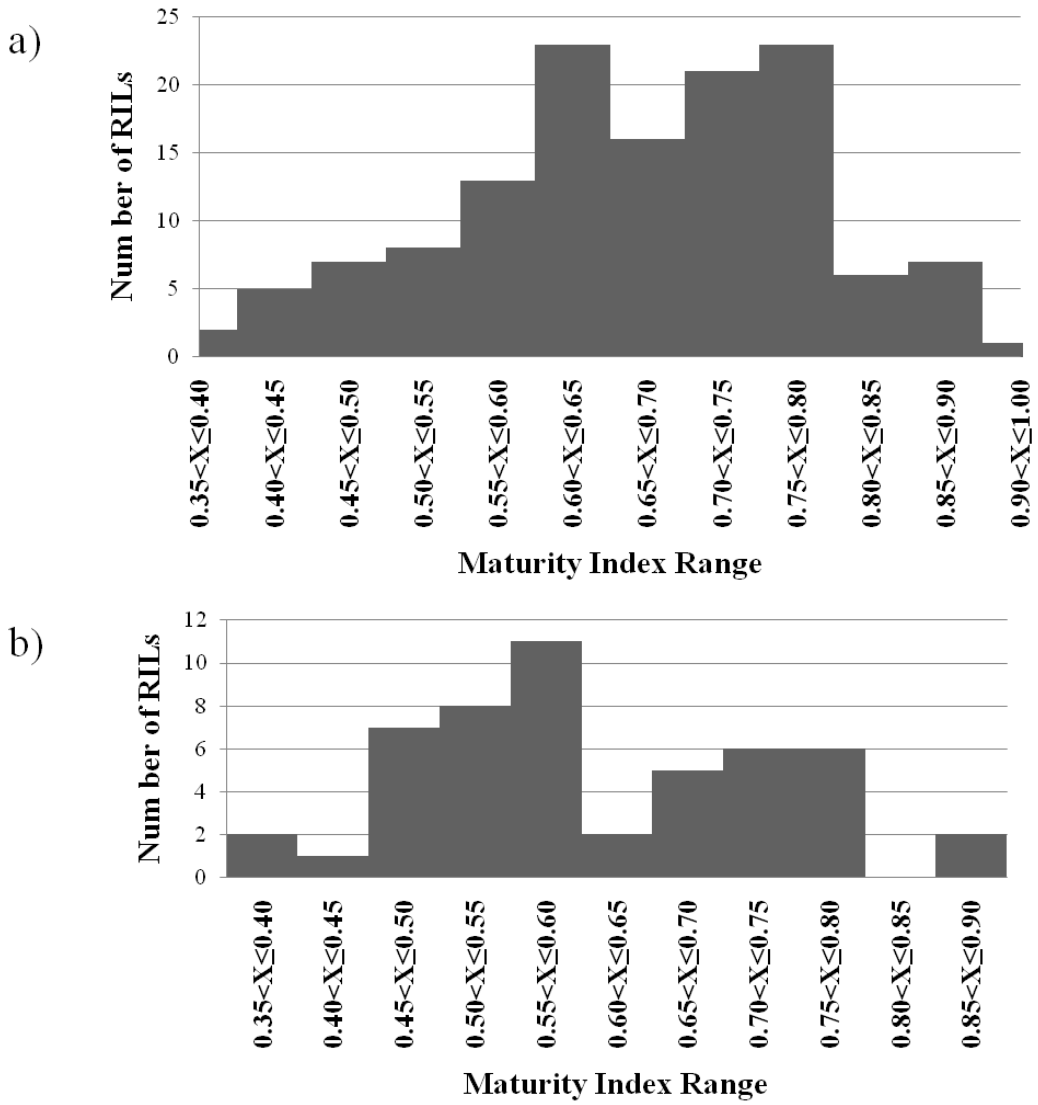

Figure 4. Maturity index distribution, groups formed each 0.05 difference in maturity index. a) Pop1 (132 RILs), b) Pop2 (50 RILs).

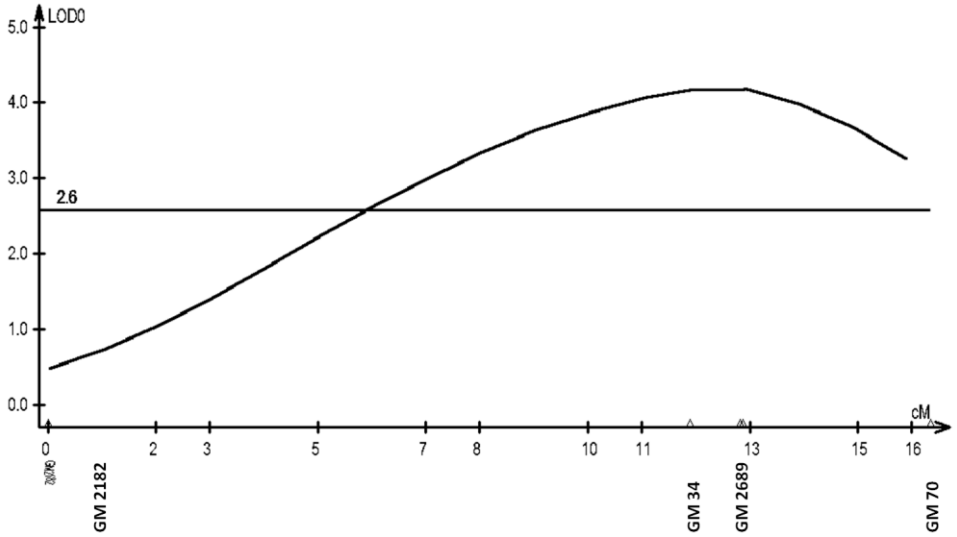

Figure 5. QTL-Maturity I (Data from Pop1). LOD score vs. map position in cM (Marker positions of LG1 plotted). Threshold to declare a QTL at 2.6 LOD- value.

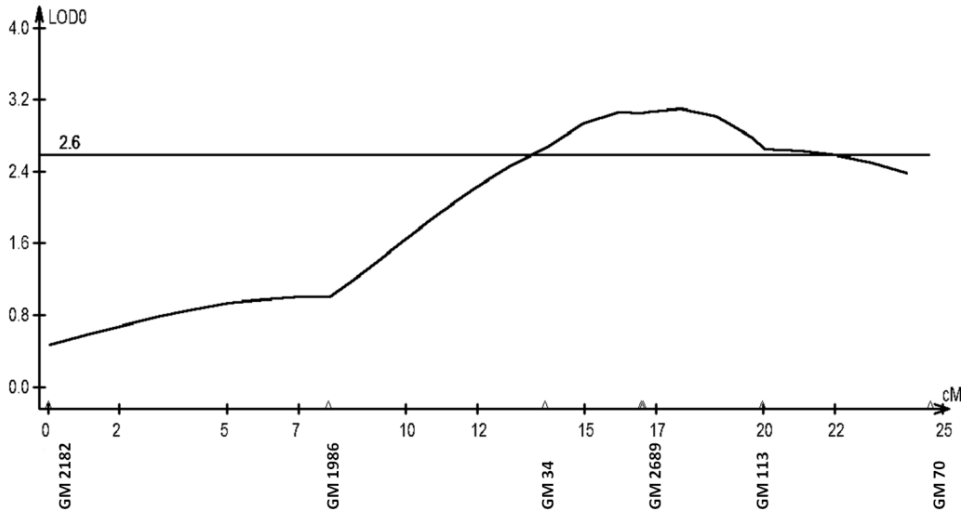

Figure 6. QTL-Maturity I (Data from Pop1 + Pop2). LOD score vs. map position in cM (Marker positions of LG1 plotted). Threshold to declare a QTL at 2.6 LOD- value.

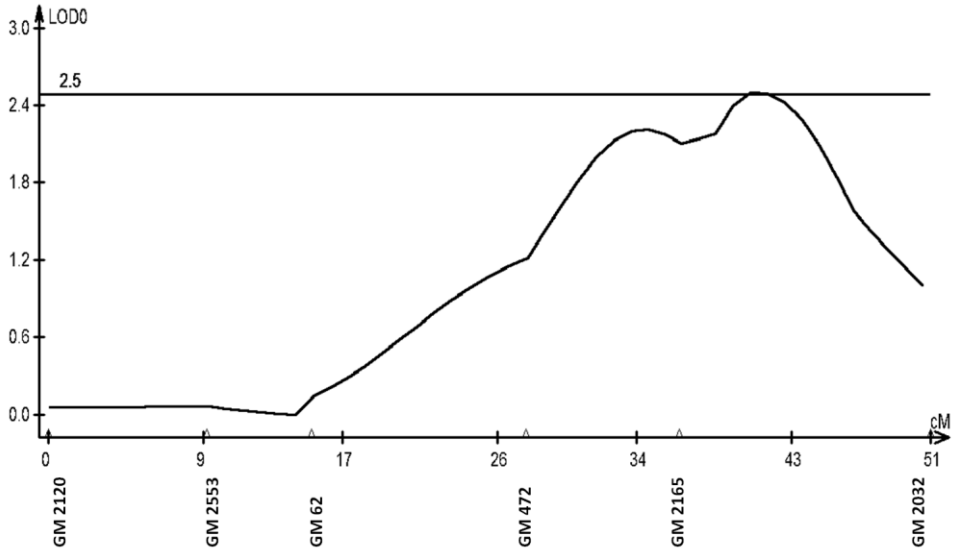

Figure 7. QTL-Maturity II (Data from Pop2). LOD score vs. map position in cM (Marker positions of LG5 plotted). Threshold to declare a QTL at 2.5 LOD- value.

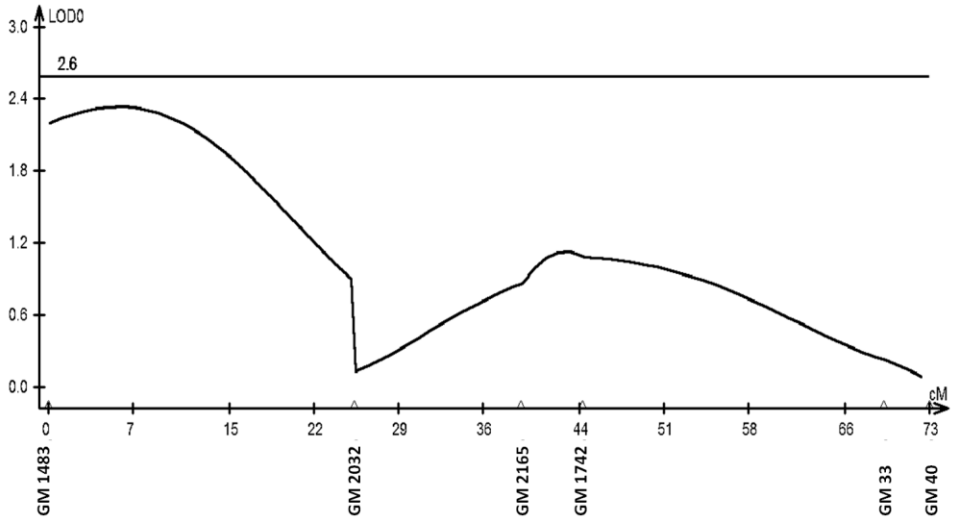

Figure 8. QTL-Maturity II (Data from Pop1, QTL is not significant in this Population). LOD score vs. map position in cM (Marker positions of LG11 plotted). Threshold to declare a QTL at 2.6 LOD- value.

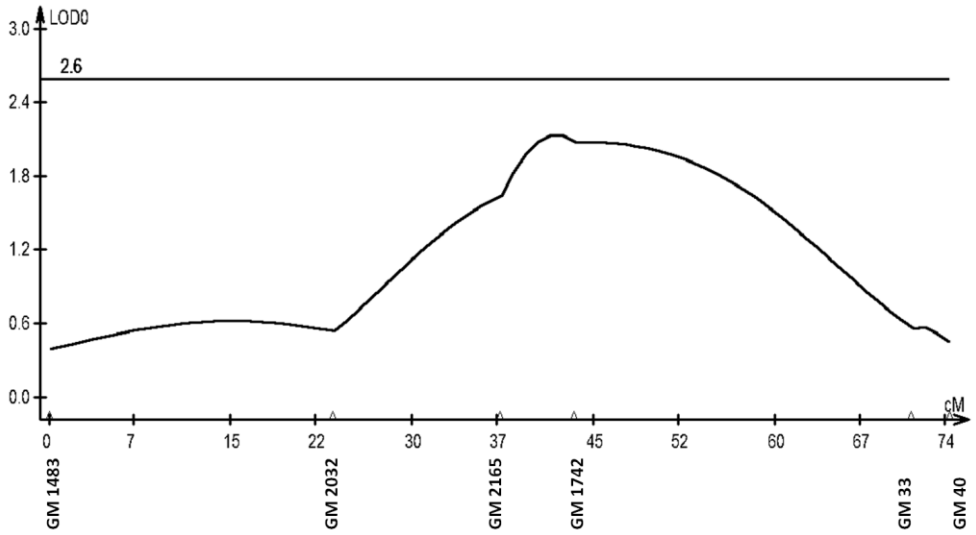

Figure 9. QTL-Maturity II (Data from Pop1 + Pop2). LOD score vs. map position in cM (Marker positions of LG12 plotted). Threshold to declare a QTL at 2.6 LOD- value.

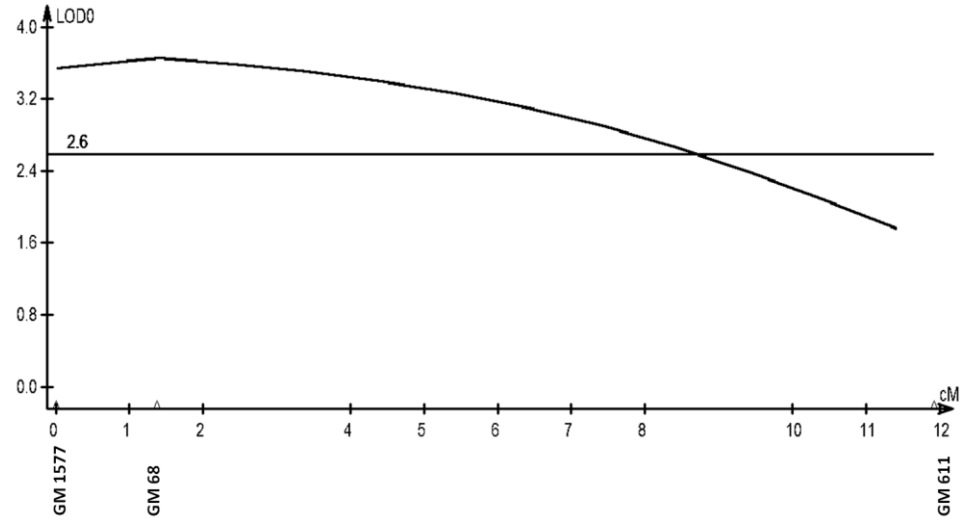

Figure 10. QTL-Maturity III(Data from Pop1+ Pop2). LOD score vs. map position in cM (Marker positions of LG11 plotted). Threshold to declare a QTL at 2.6 LOD- value.

APPENDICES

APPENDIX A. List of SSR markers used in Chapter II: Use of Simple Sequence Repeat (SSR) Markers for Mapping Quantitative Trait Loci (QTL) Influencing Early Maturity in *Arachis hypogaea* L.

Universal Name	Result parental screening	SSR Marker	Primer	Forward (5'-3')	Reverse (5'-3')
GM001	Pop1 and Pop2	Ah1TC0A01	TC0A01	CAGCTCATTTTTCACCTCCA	CCATAACCCCAAAAATGCAG
GM002	Pop1	Ah1TC1A01	TC1A01	TCAACGCGACACAAGAAGTC	GTCGGTAAATCCGACGAAAA
GM003	Pop1 and Pop2	Ah1TC1A02	TC1A02	GCAATTTGCACATTATCCGA	CATGTTCGGTTTCAAGTCTCAA
GM004	Pop1	Ah1TC1A08	TC1A08	AAGGGGTTAAGGGCATGACT	CCACAAATGGGTCGTCGAT
GM007	Pop1 and Pop2	Ah1TC1D02	TC1D02	GATCCAAAATCTCGCCTTGA	GCTGCTCTGCACAACAAGAA
GM010	Pop1 and Pop2	Ah1TC1E05	TC1E05	GAAGGATAAGCAATCGTCCA	GGATGGGATTGAACATTTGG
GM014	Pop1	Ah1TC2A02	TC2A02	CTCCCTTGTGGGTATGTGGT	GGCTCCCATTCATTCTCAAA
GM022	Pop1 and Pop2	Ah1TC2D06	TC2D06	AGGGGGAGTCAAAGGAAAGA	TCACGATCCCTTCTCCTTCA
GM023	Pop1	Ah1TC2D08	TC2D08	ATGTGGGGAGGTCGGTAAC	TCACAGGTTTTGTGTGCTCG
GM024	Pop1 and Pop2	Ah1TC2E05	TC2E05	GAATTTATAAGGCGTGGCGA	CCATCCCTTCTTCCTTCACA
GM028	Pop1 and Pop2	Ah1TC3A12	TC3A12	GCCCATATCAAGCTCCAAAA	TAGCCAGCGAAGGACTCAAT
GM032	Pop1 and Pop2	Ah1TC3E02	TC3E02	TGAAAGATAGGTTTCGGTGGA	CAAACCGAAGGAGGAACTTG
GM033	Pop1 and Pop2	Ah1TC3E05	TC3E05	CACCACTTGAGTTGGTGAGG	CTTCTTCTTCTCCCGCAATG
GM034	Pop1 and Pop2	Ah1TC3G01	TC3G01	GACGGTAATCGTGCCCTAAA	TGCAGTAGTGGCAGCAGAAC
GM038	Pop1 and Pop2	Ah1TC3H02	TC3H02	CTCTCCGCCATCCATGTAAT	ATGGTGAGCTCGACGCTAGT
GM039	Pop1 and Pop2	Ah1TC3H07	TC3H07	CAATGGGAGGCAAATCAAGT	GCCAAATGGTTCCTTCTCAA
GM040	Pop1 and Pop2	Ah1TC4A02	TC4A02	ATTCAAATCGGAATGGCAAG	GAGCAAAGGGCGAATCTATG
GM051	Pop1	Ah1TC4F02	TC4F02	GCACTGCACCCCAATCTCTA	GATGGGTGGTTTGGTGTCTC
GM056	Pop1 and Pop2	Ah1TC4G02	TC4G02	GATCCAACTGTGAATTGGGC	CACACCAGCAACAAGGAATC
GM059	Pop1 and Pop2	Ah1TC4G10	TC4G10	TTCGGTCATGTTTGTCCAGA	CTCGAGTGCTCACCCTTCAT
GM060	Pop2	Ah1TC4H02	TC4H02	ACCGCAAACTCATCCATCTC	GATAGCGTCAGAGGCAGAGG

GM062	Pop1 and Pop2	Ah1TC5A06	TC5A06	TCGGTTTGGGAGACACTCTT	TTGTAAGCAGACGCCACATC
GM066	Pop1	Ah1TC5D06	TC5D06	GAAATTTTAGTTTTCAGCACAGCA	TTTTCCCCTCTTAAATTTTCTCG
GM068	Pop1 and Pop2	Ah1TC6E01	TC6E01	CTCCCTCGCTTCCTCTTTCT	ACGCATTAACCACACACCAA
GM069	Pop1 and Pop2	Ah1TC6G09	TC6G09	GGAGGTTGCATGCATCATAGT	TCATTGAACGTATTTGAAAGCTC
GM070	Pop1 and Pop2	Ah1TC6H03	TC6H03	TCACAATCAGAGCTCCAACAA	CAGGTTCACCAGGAACGAGT
GM071	Pop1 and Pop2	Ah2TC7A02	TC7A02	CGAAAACGACACTATGAAACTGC	CCTTGGCTTACACGACTTCCT
GM072	Pop1 and Pop2	Ah2TC7C06	TC7C06	GGCAGGGGAATAAAACTACTAACT	TTTTCCTTCCTTCTCCTTTGTC
GM074	Pop1 and Pop2	Ah2TC7E04	TC7E04	GAAGGACCCCATCTATTCAAA	TCCGATTTCTCTCTCTCTCTC
GM076	Pop1 and Pop2	Ah2TC7G10	TC7G10	AATGGGGTTCACAAGAGAGAGA	CCAGCCATGCACTCATAGAATA
GM079	Pop1	Ah2TC7H11	TC7H11	AGGTTGGAACTATGGCTGATTG	CCAGTTTAGCATGTGTGGTTCA
GM081	Pop1 and Pop2	Ah2TC9B08	TC9B08	GGTTGGGTTGAGAACAAGG	ACCCTCACCACTAACTCCATTA
GM083	Pop1 and Pop2	Ah2TC9C06	TC9C06	CAAATGGCAGAGTGCGTCTA	CCCTCCTGACTGGGTCCT
GM084	Pop1 and Pop2	Ah2TC9C08	TC9C08	ACTTTTGGGGCAGGATGAG	GCCTCTATTGCTGAGATTATTGC
GM089	Pop1	Ah2TC9H08	TC9H08	GCCAAAGGGGACCATAAAC	TCCATCTTCCATCTCATCCAC
GM092	Pop1 and Pop2	Ah2TC11A04	TC11A04	ACTCTGCATGGATGGCTACAG	CATGTTCGGTTTCAAGTCTCAA
GM093	Pop1 and Pop2	Ah2TC11B04	TC11B04	GATCTGAAGGCTCTGATACCAT	GATCTCAACCAGAACAGTATGC
GM097	Pop1 and Pop2	Ah2TC11E04	TC11E04	ACGACACCCTGAAATCAAGTTT	CCGAAGGCACCAAAAAGTAT
GM099	Pop1 and Pop2	Ah2TC11H06	TC11H06	CCATGTGAGGTATCAGTAAAGAAAGG	CCACCAACAACATTGGATGAAT
GM113	Pop1 and Pop2	Ah2AC2C12	AC2C12	TATCGAGCCGAATATGAAT	GCAGGATTTTGTAATTGAGAG
GM118	Pop1 and Pop2	Ah2AC3C07	AC3C07	GGGGGTTTAGGAGCAAGATTT	CAAGGTGAGAACAAAGGCAAAG
GM126	Pop1 and Pop2	AS1RI1F06	RI1F06	TGTCTCTCTTCCTTTCCTTGCT	CCTTTTGCTTCTTTGCTTCC
GM228	Pop1 and Pop2	gi-951107	gi-1107	GATACATCTCATCCGTTCGTG	CCGTCCGACCACATACAA
GM230	Pop2	gi-28194660	gi-4660	GGCGAACATGGCCACGAC	TGACCCCATGCACCTTGACATA
GM245	Pop1 and Pop2	pPGPSeq3E5	Seq3E05	CGATGAGGACAGAGACACGA	CGCTTGAACCCGACTATTTT
GM254	Pop1 and Pop2	pPGPSeq4F10	Seq4F10	TGCGAAACCCCTAACTGACT	TCTATGTTGCTGCCGTTGAC
GM266	Pop1 and Pop2	pPGSSeq14G3	Seq14G03	CAATTTTATTTGCCACATGCT	TCGCAGTTCTCAAAGTTATCG

GM338	Pop1 and Pop2	Ah-745	Ah-745	TGTTGTTCTGCTCCTGCTTTTG	ATTCGGACCAAAATGTCCCTTC
GM339	Pop1 and Pop2	Ah4-04	Ah 4-4	CGATTTCTTTACTGAGTGAG	ATTTTTTTGCTCCACACA
GM342	Pop1 and Pop2	Ah4-26	Ah 4-26	TGGAATCTATTGCTCATCGGCTCTG	CTCACCCATCATCATCGTCACATT
GM344	Pop1 and Pop2	Lec-1	Lec1	CAAGCATCAACAACAACGA	GTCCGACCACATACAAGAGTT
GM346	Pop1 and Pop2	PM179	PM-179	CTGATGCATGTTTAGCACACTT	TGAGTTGTGACGGCTTGTGT
GM367	Pop1 and Pop2	PM3	PM3	GAAAGAAATTATACACTCCAATTATGC	CGGCATGACAGCTCTATGTT
GM371	Pop1 and Pop2	PM15	PM15	CCTTTTCTAACACATTCACACATGA	GGCTCCCTTCGATGATGAC
GM377	Pop1 and Pop2	PM35	PM35	TGTGAAACCAAATCACTTTCATTC	TGGTGAAAAGAAAGGGGAAA
GM378	Pop1 and Pop2	PM36	PM36	ACTCGCCATAGCCAACAAAC	CATTCCCACAACTCCCACAT
GM381	Pop1 and Pop2	PM42	PM42	ACGGGCCAAGTGAAGTGAT	TCTTGCTTCTTTGGTGATTAGC
GM382	Pop1 and Pop2	PM45	PM45	TGAGTTGTGACGGCTTGTGT	GATGCATGTTTAGCACACTTGA
GM393	Pop1 and Pop2	PM137	PM137	AACCAATTCAACAAACCCAGT	GAAGATGGATGAAAACGGATG
GM404	Pop1 and Pop2	PM201	PM201	CCTTTATAGAGGACCTTCCCTCTC	GCCTATTTGGTATCGGCTCA
GM405	Pop1 and Pop2	PM204	PM204	TGGGCCTAAACCCAACCTAT	CCACAAACAGTGCAGCAATC
GM408	Pop1	PM210	PM210	CCGCAGATCTTCTCCTGTGT	CCTCCTCATCCTCTAAACTCTGC
GM414	Pop1	PM230	PM230	GCATTGCTCAATGATGAATAACA	ATCTTCCACACCGCCATTT
GM415	Pop2	PM238	PM238	CTCTCCTCTGCTCTGCACTG	ACAAGAACATGGGGATGAAGA
GM422	Pop1 and Pop2	pPGPseq2A5		GGGAATAGCGAGATACATGTCAG	CAGGAGAGAAGGATTGTGCC
GM426	Pop1 and Pop2	pPGPseq2C10		GCAAGTCACATAGTTCAATTTTGG	GGCATAGCCATCCAAATCAT
GM429	Pop1 and Pop2	pPGPseq2E6		TACAGCATTGCCTTCTGGTG	CCTGGGCTGGGGTATTATTT
GM431	Pop2	pPGPseq2F5		TGACCAAAGTGATGAAGGGA	AAGTTGTTTGTACATCTGTCATCG
GM434	Pop1 and Pop2	pPGPseq2G4		TTCTTGGTTCCTTTGGCTTC	TGCTCAAGTGTCCTTATTGGTG
GM437	Pop1 and Pop2	pPGPseq3A1		ATCATTGTGCTGAGGGAAGG	CACCATTTTTCTTTTTCACCG
GM439	Pop1 and Pop2	pPGPseq3A6		TGCATCAGCAAGCTACATACG	GCGATTCACCATCAATCTCA
GM440	Pop1 and Pop2	pPGPseq3A8		ATACGTGACTTGGGCCAGAC	AGTGAAAAATACACCCAACGAA
GM445	Pop1 and Pop2	pPGPseq3B7		GGCTCGTCAATCAGAACTCC	TGGGTTAAACGGGCTAAAAA

GM446	Pop1 and Pop2	pPGPseq3B8		GGAGAAAGATCAAACGAGAACA	TTCGAATATCTGACATTTGCTTTT
GM450	Pop1 and Pop2	pPGPseq3E10		TCCCAAAAATAACAAACATGGA	ACGCTTTGAGACTCGTCGTT
GM455	Pop1 and Pop2	pPGPseq3H6		CGCAGGCTATAACTAGCTCCC	CATAAAGCAAATGGCGACAA
GM464	Pop1 and Pop2	pPGPseq4E10		TTCTTCCATGTTTCCTGATGG	AATGCTTGCAACGGATACCT
GM465	Pop1 and Pop2	pPGPseq4E12		TTTTCTTAACACCCTACCCCC	CCAGCTCATGCAAACAAACA
GM468	Pop1 and Pop2	pPGPseq4F9		ACGTGAAATCTGGCTGGAAA	ACAATCCACACGCCAACATA
GM472	Pop1 and Pop2	pPGPseq4H11		ATCACCATCAGAACGATCCC	TTTGTAGCCTTCTGGCGAGT
GM477	Pop2	pPGPseq5D5		AAAAGAAAGACCTTCCCCGA	GCAGGTAATCTGCCGTGATT
GM488	Pop1 and Pop2	pPGPseq7G2		ACTCCCGATGCACTTGAAAT	AACCTCTGTGCACTGTCCCT
GM489	Pop1 and Pop2	pPGPseq7H6		CATCCTCACGGGAGTCAGAT	ATACCTACGCGTTGTGGAGC
GM496	Pop1 and Pop2	pPGPseq8E12		TCTGTTGAGAACCACCAGCA	GTGCTAGTTGCTTGACGCAC
GM507	Pop2	pPGSseq11G3		CCGCGTTGTTAAACCAGAAC	ATGGAGGATGTGAGTGGGAA
GM508	Pop1 and Pop2	pPGSseq11G7		CATGTCTCCATGAGCATTTCA	TGGATGTGGACAGCATATCG
GM514	Pop1 and Pop2	pPGSseq12E10		TGCTTTTAGAGGCTTTGCCA	GAAACTGCAACAGCAACAGAA
GM516	Pop1 and Pop2	pPGSseq12F7		TGTCGTTGTAAGACCTCGGA	TTGGTTTCCTTAAGGCTTCG
GM518	Pop1 and Pop2	pPGSseq13A10		AACTCGCTTGTACCGGCTAA	AGGAATAATAACAATACCAACAGCA
GM525	Pop1 and Pop2	pPGSseq13E6A		TGGCAATTTATTGATGCAGG	GTCACGTAATTGGATGCACG
GM534	Pop1 and Pop2	pPGSseq14F4		ACGTTTAGTTGCTTGCGTGA	TGAATTCAAAGGAAAATGAAAAA
GM541	Pop1	pPGSseq15C12		ACAATGCAATGACCGTTGTT	TTGTTGCATGAGAACGTGAA
GM544	Pop1 and Pop2	pPGSseq15D3		CATGCCATCATCACAACACA	GGAGGAAGCAATGGTTTCAG
GM548	Pop1 and Pop2	pPGSseq15F12		AAAGTCAACCGCTCACACTG	AGGGTTAGGATTTTGGGTGG
GM556	Pop1 and Pop2	pPGSseq16F1		TGCTTCCATCAGCTTTTCCT	AAATGAGGGCCTCCAAAGTT
GM557	Pop1 and Pop2	pPGSseq16F10		TGGAGGGAAAAACATTTTGG	CCTGGAGGGGTGAGAGGT
GM560	Pop1 and Pop2	pPGSseq16G8		CTCAAAAAGCGCTTAGCCAC	CTGCCTACTGCCTACTGCCT
GM567	Pop1 and Pop2	pPGSseq17F6		CGTCGGATTTATCTGCCAGT	AGTAGGGGCAAGGGTTGATG
GM571	Pop1	pPGSseq18A5		TGATTCGATTTACTCATGCACA	GAGGATTCTTGAGCCTCGAC

GM577	Pop1 and Pop2	pPGSseq18C5		GGACAGCCGGATGCTATTTA	ACATGAGTCCCTTTTCCCTT
GM580	Pop1 and Pop2	pPGSseq18G1		AATAGGTTGTGAAGCACGCA	TTCGGTGGTACTTTTAAGGCA
GM590	Pop1 and Pop2	pPGSseq19C3		TCATCGCCAAACTCTTCTCC	TCGAAGAGTGCATGTTGACC
GM594	Pop1	pPGSseq19E9		ACTGCTTGCTCTCTTCCTCG	TTCCACCTATAAAATCAATGGTGA
GM600	Pop1 and Pop2	pPGSseq19G7		ATTCAATTCCTCTCTCCCCC	TCAATCAATCAATCGCAGGA
GM611	Pop1 and Pop2	pPGSseq9G5		CAAATTGTGCAGCCAAGAGA	CATATGCCCAGGAAGAGGAA
GM613	Pop1 and Pop2	GA1		GCGTGAAATGAGTGTTGTGAG	CATAGCCACCATAGACACCAAA
GM618	Pop1 and Pop2	GA8		TGAAGGGAGGAAGAGACGAAT	CCTCTGTGCGACATGACTCTAA
GM623	Pop1 and Pop2	GA21		CAGGATGAACAGGCACAGAAT	ATGAACAATTGCGATTTGGAC
GM624	Pop1 and Pop2	GA24		AACGAAATATTTTGAGAAAGGAT	AGCATTAGCAACTCTAAGCTCAT
GM627	Pop1 and Pop2	GA28		AGATGGTGGTGTAGGAGTTGTGT	TGGCCGTTGGATATTTATTTG
GM630	Pop1 and Pop2	GA32		CAGCAATTCAGCAAACTAATGAA	TCCTCCCACGTCCTTTTATTT
GM631	Pop1 and Pop2	GA33		CAAGTGATAGCACGCTGTTTG	TTAAGTCCCATGCCTGTCTTG
GM633	Pop1	GA35		CAAAGTTTGCAGTGATTTTGTTG	AAATTTTCAGGTAAATCATTCTT
GM637	Pop1	GA44		TGACTTTAATTTTGAGCTTCCTATAA	TTTTCTGTCCATAATTATATCGTATTT
GM641	Pop1 and Pop2	GA49		ACGTTTCCCCAATAAGACCAC	TGGACACCTTATCGGCTTATC
GM657	Pop1 and Pop2	GA72		ACTTTGGTGGCTTTCCTTCAT	TCTCTGTGCCCTCTTTCTTCA
GM660	Pop1 and Pop2	GA80		TGAAAGTAACTCGTTTACAGTTTGAAG	TCACTAAACATGTGGGTAACTAAGAAA
GM679	Pop1 and Pop2	GA133		GGTGTTATGTATAGCCACCAG	AAATAGTATGGACCAGAAATAATAAG
GM682	Pop1	GA140		TTAGGCTGGTGGAAAGTGATG	CAAATAAAACAATGAATTGATAATCG
GM690	Pop2	GA156		CTACTCCCTCTGCTGCTTCCT	TAGGGTTTCGTTGAGGAGGTT
GM691	Pop1	GA160		TCTTTATCCCGATGAATGAAA	CTCCCACAAACACAAACACAC
GM692	Pop1 and Pop2	GA161		TGAGGCCGTCTTGTTTAGAGA	CCTCTTCCATCACCGTTCATA
GM693	Pop1 and Pop2	GA163		ATGTATAGTGGCGGATCCAAT	TTTTGAAGTATTCTCTTTTTCAACA
GM695	Pop1 and Pop2	GA166		TGGTCGCAGATAGTATTTCTCCT	TGGAATTTGAATCGCACTCTT

APPENDIX B. SAS program used to analyze phenotypic data.

```
options linesize=240 pagesize=200 ;

data a ; infile "C:\Documents and Settings\fvilleg\My Documents\Writing_Your favorite part!\Heritability paper\Phenotypic Data\All Phenotypic data 08 EMS.txt";
input plot 1-4 rep 6 block 8-9 entry 11-13 type $ & 21-28 F2 29-30 F6 32-33 maturity 35-40 Yield 42-45 PodLength 47-50 PodWidth 52-55 PodWeight 57-60 Noseeds 62-65 SeedLength 67-70 SeedWidth 72-75 SeedWeight 77-80 LScore 82-86 aScore 88-91 bScore 93-97 ;
f21=0 ; f61=0 ; if type='Cross1' then do ; f21=f2 ; f61=f6 ; end ;
f22=0 ; f62=0 ; if type='Cross2' then do ; f22=f2 ; f62=f6 ; end ;
geno=00 ; if type='Check' then geno=entry ;
year=2008 ;

data b ; infile "C:\Documents and Settings\fvilleg\My Documents\Writing_Your favorite part!\Heritability paper\Phenotypic Data\All Phenotypic data 09 EMS.txt";
input plot 1-4 rep 6 block 8-9 entry 11-13 type $ & 21-28 F2 29-30 F6 32-33 maturity 35-40 Yield 42-45 PodLength 47-50 PodWidth 52-55 PodWeight 57-60 Noseeds 62-65 SeedLength 67-70 SeedWidth 72-75 SeedWeight 77-80 LScore 82-86 aScore 88-91 bScore 93-97 ;
f21=0 ; f61=0 ; if type='Cross1' then do ; f21=f2 ; f61=f6 ; end ;
f22=0 ; f62=0 ; if type='Cross2' then do ; f22=f2 ; f62=f6 ; end ;
geno=00 ; if type='Check' then geno=entry ;
year=2009 ;

data a ; set a b ;

proc print data=a ; var year plot--geno;
proc means data=a ; var maturity Yield PodLength PodWidth PodWeight Noseeds SeedLength SeedWidth SeedWeight LScore aScore bScore ;
run ;

proc mixed data=a covtest ; class year rep block type f21 f61 f22 f62 geno ;
```

```
model maturity = type ;
random year rep(year) block(year rep)
            f21(type) f22(type) f61(type f21) f62(type f22) geno(type)
      year*type year*f21(type) year*f22(type) year*f61(type f21) year*f62(type f22) year*geno(type) ;
lsmeans type / pdiff ;
estimate 'Female1 vs Male'          type    0     0     0     1     0     -1     ;
estimate 'Cross1 vs Mid-parent1'    type    0     1     0   -0.5    0   -0.5     ;
estimate 'Female2 vs Male'          type    0     0     0     0     1     -1     ;
estimate 'Cross2 vs Mid-parent2'    type    0     0     1     0   -0.5   -0.5     ;
run ;

proc mixed data=a covtest ; class year rep block type f21 f61 f22 f62 geno ;
model Yield = type ;
random year rep(year) block(year rep)
            f21(type) f22(type) f61(type f21) f62(type f22) geno(type)
      year*type year*f21(type) year*f22(type) year*f61(type f21) year*f62(type f22) year*geno(type) ;
lsmeans type / pdiff ;
estimate 'Female1 vs Male'          type    0     0     0     1     0     -1     ;
estimate 'Cross1 vs Mid-parent1'    type    0     1     0   -0.5    0   -0.5     ;
estimate 'Female2 vs Male'          type    0     0     0     0     1     -1     ;
estimate 'Cross2 vs Mid-parent2'    type    0     0     1     0   -0.5   -0.5     ;

run ;

proc mixed data=a covtest ; class year rep block type f21 f61 f22 f62 geno ;
model PodLength = type ;
random year rep(year) block(year rep)
            f21(type) f22(type) f61(type f21) f62(type f22) geno(type)
      year*type year*f21(type) year*f22(type) year*f61(type f21) year*f62(type f22) year*geno(type) ;
lsmeans type / pdiff ;
estimate 'Female1 vs Male'          type    0     0     0     1     0     -1     ;
estimate 'Cross1 vs Mid-parent1'    type    0     1     0   -0.5    0   -0.5     ;
estimate 'Female2 vs Male'          type    0     0     0     0     1     -1     ;
```

```
estimate 'Cross2 vs Mid-parent2'        type      0         0         1         0       -0.5       -0.5          ;
run ;

proc mixed data=a covtest ; class year rep block type f21 f61 f22 f62 geno ;
model PodWidth = type ;
random year rep(year) block(year rep)
            f21(type) f22(type) f61(type f21) f62(type f22) geno(type)
      year*type year*f21(type) year*f22(type) year*f61(type f21) year*f62(type f22) year*geno(type) ;
lsmeans type / pdiff ;
estimate 'Female1 vs Male'              type      0         0         0         1         0         -1          ;
estimate 'Cross1 vs Mid-parent1'        type      0         1         0       -0.5        0       -0.5          ;
estimate 'Female2 vs Male'              type      0         0         0         0         1         -1          ;
estimate 'Cross2 vs Mid-parent2'        type      0         0         1         0       -0.5       -0.5          ;
run ;

proc mixed data=a covtest ; class year rep block type f21 f61 f22 f62 geno ;
model PodWeight = type ;
random year rep(year) block(year rep)
            f21(type) f22(type) f61(type f21) f62(type f22) geno(type)
      year*type year*f21(type) year*f22(type) year*f61(type f21) year*f62(type f22) year*geno(type) ;
lsmeans type / pdiff ;
estimate 'Female1 vs Male'              type      0         0         0         1         0         -1          ;
estimate 'Cross1 vs Mid-parent1'        type      0         1         0       -0.5        0       -0.5          ;
estimate 'Female2 vs Male'              type      0         0         0         0         1         -1          ;
estimate 'Cross2 vs Mid-parent2'        type      0         0         1         0       -0.5       -0.5          ;
run ;

proc mixed data=a covtest ; class year rep block type f21 f61 f22 f62 geno ;
model Noseeds = type ;
random year rep(year) block(year rep)
            f21(type) f22(type) f61(type f21) f62(type f22) geno(type)
      year*type year*f21(type) year*f22(type) year*f61(type f21) year*f62(type f22) year*geno(type) ;
lsmeans type / pdiff ;
```

```
estimate 'Female1 vs Male'          type    0     0     0     1     0     -1     ;
estimate 'Cross1 vs Mid-parent1'    type    0     1     0   -0.5    0   -0.5     ;
estimate 'Female2 vs Male'          type    0     0     0     0     1     -1     ;
estimate 'Cross2 vs Mid-parent2'    type    0     0     1     0   -0.5   -0.5     ;
run ;

proc mixed data=a covtest ; class year rep block type f21 f61 f22 f62 geno ;
model SeedLength = type ;
random year rep(year) block(year rep)
            f21(type) f22(type) f61(type f21) f62(type f22) geno(type)
      year*type year*f21(type) year*f22(type) year*f61(type f21) year*f62(type f22) year*geno(type) ;
lsmeans type / pdiff ;
estimate 'Female1 vs Male'          type    0     0     0     1     0     -1     ;
estimate 'Cross1 vs Mid-parent1'    type    0     1     0   -0.5    0   -0.5     ;
estimate 'Female2 vs Male'          type    0     0     0     0     1     -1     ;
estimate 'Cross2 vs Mid-parent2'    type    0     0     1     0   -0.5   -0.5     ;
run ;

proc mixed data=a covtest ; class year rep block type f21 f61 f22 f62 geno ;
model SeedWidth = type ;
random year rep(year) block(year rep)
            f21(type) f22(type) f61(type f21) f62(type f22) geno(type)
      year*type year*f21(type) year*f22(type) year*f61(type f21) year*f62(type f22) year*geno(type) ;
lsmeans type / pdiff ;
estimate 'Female1 vs Male'          type    0     0     0     1     0     -1     ;
estimate 'Cross1 vs Mid-parent1'    type    0     1     0   -0.5    0   -0.5     ;
estimate 'Female2 vs Male'          type    0     0     0     0     1     -1     ;
estimate 'Cross2 vs Mid-parent2'    type    0     0     1     0   -0.5   -0.5     ;
run ;

proc mixed data=a covtest ; class year rep block type f21 f61 f22 f62 geno ;
model SeedWeight = type ;
random year rep(year) block(year rep)
```

```
            f21(type) f22(type) f61(type f21) f62(type f22) geno(type)
      year*type year*f21(type) year*f22(type) year*f61(type f21) year*f62(type f22) year*geno(type) ;
lsmeans type / pdiff ;
estimate 'Female1 vs Male'          type    0     0     0     1     0     -1     ;
estimate 'Cross1 vs Mid-parent1'    type    0     1     0   -0.5    0   -0.5     ;
estimate 'Female2 vs Male'          type    0     0     0     0     1     -1     ;
estimate 'Cross2 vs Mid-parent2'    type    0     0     1     0   -0.5   -0.5     ;
run ;

proc mixed data=a covtest ; class year rep block type f21 f61 f22 f62 geno ;
model LScore = type ;
random year rep(year) block(year rep)
            f21(type) f22(type) f61(type f21) f62(type f22) geno(type)
      year*type year*f21(type) year*f22(type) year*f61(type f21) year*f62(type f22) year*geno(type) ;
lsmeans type / pdiff ;
estimate 'Female1 vs Male'          type    0     0     0     1     0     -1     ;
estimate 'Cross1 vs Mid-parent1'    type    0     1     0   -0.5    0   -0.5     ;
estimate 'Female2 vs Male'          type    0     0     0     0     1     -1     ;
estimate 'Cross2 vs Mid-parent2'    type    0     0     1     0   -0.5   -0.5     ;
run ;

proc mixed data=a covtest ; class year rep block type f21 f61 f22 f62 geno ;
model aScore = type ;
random year rep(year) block(year rep)
            f21(type) f22(type) f61(type f21) f62(type f22) geno(type)
      year*type year*f21(type) year*f22(type) year*f61(type f21) year*f62(type f22) year*geno(type) ;
lsmeans type / pdiff ;
estimate 'Female1 vs Male'          type    0     0     0     1     0     -1     ;
estimate 'Cross1 vs Mid-parent1'    type    0     1     0   -0.5    0   -0.5     ;
estimate 'Female2 vs Male'          type    0     0     0     0     1     -1     ;
estimate 'Cross2 vs Mid-parent2'    type    0     0     1     0   -0.5   -0.5     ;
run ;
```

```
proc mixed data=a covtest ; class year rep block type f21 f61 f22 f62 geno ;
model bScore = type ;
random year rep(year) block(year rep)
          f21(type) f22(type) f61(type f21) f62(type f22) geno(type)
     year*type year*f21(type) year*f22(type) year*f61(type f21) year*f62(type f22) year*geno(type) ;
lsmeans type / pdiff ;
estimate 'Female1 vs Male'          type    0     0     0      1      0      -1      ;
estimate 'Cross1 vs Mid-parent1'    type    0     1     0    -0.5     0    -0.5      ;
estimate 'Female2 vs Male'          type    0     0     0      0      1      -1      ;
estimate 'Cross2 vs Mid-parent2'    type    0     0     1      0    -0.5   -0.5      ;
run ;

*/Proc glm, in order to obtain the correct test for the contrasts*/

proc glm data=a ; class entry year rep block type ;
model Maturity = year rep(year) block(year rep) type entry(type) year*entry(type) ;
lsmeans type / stderr pdiff e=year*entry(type) ;
contrast 'Female1 vs Male'          type    0     0     0      1      0      -1      /e=year*entry(type) ;
contrast 'Cross1 vs Mid-parent1'    type    0     1     0    -0.5     0    -0.5      /e=year*entry(type) ;
contrast 'Female2 vs Male'          type    0     0     0      0      1      -1      /e=year*entry(type) ;
contrast 'Cross2 vs Mid-parent2'    type    0     0     1      0    -0.5   -0.5      /e=year*entry(type) ;
contrast 'Cross1 vs Female1'        type    0     1     0     -1      0      0       /e=year*entry(type) ;
contrast 'Cross2 vs Female2'        type    0     0     1      0     -1      0       /e=year*entry(type) ;
contrast 'Female1 vs Female2'       type    0     0     0      1     -1      0       /e=year*entry(type) ;
estimate 'Female1 vs Male'          type    0     0     0      1      0      -1      ;
estimate 'Cross1 vs Mid-parent1'    type    0     1     0    -0.5     0    -0.5      ;
estimate 'Female2 vs Male'          type    0     0     0      0      1      -1      ;
estimate 'Cross2 vs Mid-parent2'    type    0     0     1      0    -0.5   -0.5      ;
estimate 'Cross1 vs Female1'        type    0     1     0     -1      0      0       ;
estimate 'Cross2 vs Female2'        type    0     0     1      0     -1      0       ;
run ;

proc glm data=a ; class entry year rep block type ;
```

```
model Yield = year rep(year) block(year rep) type entry(type) year*entry(type) ;
lsmeans type / stderr pdiff e=year*entry(type) ;
contrast 'Female1 vs Male'          type  0   0   0    1    0    -1   /e=year*entry(type) ;
contrast 'Cross1 vs Mid-parent1'    type  0   1   0  -0.5    0  -0.5   /e=year*entry(type) ;
contrast 'Female2 vs Male'          type  0   0   0    0    1    -1   /e=year*entry(type) ;
contrast 'Cross2 vs Mid-parent2'    type  0   0   1    0  -0.5  -0.5   /e=year*entry(type) ;
contrast 'Cross1 vs Female1'        type  0   1   0   -1    0    0    /e=year*entry(type) ;
contrast 'Cross2 vs Female2'        type  0   0   1    0   -1    0    /e=year*entry(type) ;
contrast 'Female1 vs Female2'       type  0   0   0    1   -1    0    /e=year*entry(type) ;
estimate 'Female1 vs Male'          type  0   0   0    1    0    -1   ;
estimate 'Cross1 vs Mid-parent1'    type  0   1   0  -0.5    0  -0.5   ;
estimate 'Female2 vs Male'          type  0   0   0    0    1    -1   ;
estimate 'Cross2 vs Mid-parent2'    type  0   0   1    0  -0.5  -0.5   ;
estimate 'Cross1 vs Female1'        type  0   1   0   -1    0    0    ;
estimate 'Cross2 vs Female2'        type  0   0   1    0   -1    0    ;
run ;

proc glm data=a ; class entry year rep block type ;
model PodLength = year rep(year) block(year rep) type entry(type) year*entry(type) ;
lsmeans type / stderr pdiff e=year*entry(type) ;
contrast 'Female1 vs Male'          type  0   0   0    1    0    -1   /e=year*entry(type) ;
contrast 'Cross1 vs Mid-parent1'    type  0   1   0  -0.5    0  -0.5   /e=year*entry(type) ;
contrast 'Female2 vs Male'          type  0   0   0    0    1    -1   /e=year*entry(type) ;
contrast 'Cross2 vs Mid-parent2'    type  0   0   1    0  -0.5  -0.5   /e=year*entry(type) ;
contrast 'Cross1 vs Female1'        type  0   1   0   -1    0    0    /e=year*entry(type) ;
contrast 'Cross2 vs Female2'        type  0   0   1    0   -1    0    /e=year*entry(type) ;
contrast 'Female1 vs Female2'       type  0   0   0    1   -1    0    /e=year*entry(type) ;
estimate 'Female1 vs Male'          type  0   0   0    1    0    -1   ;
estimate 'Cross1 vs Mid-parent1'    type  0   1   0  -0.5    0  -0.5   ;
estimate 'Female2 vs Male'          type  0   0   0    0    1    -1   ;
estimate 'Cross2 vs Mid-parent2'    type  0   0   1    0  -0.5  -0.5   ;
estimate 'Cross1 vs Female1'        type  0   1   0   -1    0    0    ;
estimate 'Cross2 vs Female2'        type  0   0   1    0   -1    0    ;
```

```
run ;

proc glm data=a ; class entry year rep block type ;
model PodWidth = year rep(year) block(year rep) type entry(type) year*entry(type) ;
lsmeans type / stderr pdiff e=year*entry(type) ;
contrast 'Female1 vs Male'          type  0  0  0   1    0   -1   /e=year*entry(type) ;
contrast 'Cross1 vs Mid-parent1'    type  0  1  0  -0.5  0  -0.5  /e=year*entry(type) ;
contrast 'Female2 vs Male'          type  0  0  0   0    1   -1   /e=year*entry(type) ;
contrast 'Cross2 vs Mid-parent2'    type  0  0  1   0   -0.5 -0.5  /e=year*entry(type) ;
contrast 'Cross1 vs Female1'        type  0  1  0  -1    0    0   /e=year*entry(type) ;
contrast 'Cross2 vs Female2'        type  0  0  1   0   -1   0    /e=year*entry(type) ;
contrast 'Female1 vs Female2'       type  0  0  0   1   -1   0    /e=year*entry(type) ;
estimate 'Female1 vs Male'          type  0  0  0   1    0   -1   ;
estimate 'Cross1 vs Mid-parent1'    type  0  1  0  -0.5  0  -0.5  ;
estimate 'Female2 vs Male'          type  0  0  0   0    1   -1   ;
estimate 'Cross2 vs Mid-parent2'    type  0  0  1   0   -0.5 -0.5  ;
estimate 'Cross1 vs Female1'        type  0  1  0  -1    0    0   ;
estimate 'Cross2 vs Female2'        type  0  0  1   0   -1   0    ;
run ;

proc glm data=a ; class entry year rep block type ;
model PodWeight = year rep(year) block(year rep) type entry(type) year*entry(type) ;
lsmeans type / stderr pdiff e=year*entry(type) ;
contrast 'Female1 vs Male'          type  0  0  0   1    0   -1   /e=year*entry(type) ;
contrast 'Cross1 vs Mid-parent1'    type  0  1  0  -0.5  0  -0.5  /e=year*entry(type) ;
contrast 'Female2 vs Male'          type  0  0  0   0    1   -1   /e=year*entry(type) ;
contrast 'Cross2 vs Mid-parent2'    type  0  0  1   0   -0.5 -0.5  /e=year*entry(type) ;
contrast 'Cross1 vs Female1'        type  0  1  0  -1    0    0   /e=year*entry(type) ;
contrast 'Cross2 vs Female2'        type  0  0  1   0   -1   0    /e=year*entry(type) ;
contrast 'Female1 vs Female2'       type  0  0  0   1   -1   0    /e=year*entry(type) ;
estimate 'Female1 vs Male'          type  0  0  0   1    0   -1   ;
estimate 'Cross1 vs Mid-parent1'    type  0  1  0  -0.5  0  -0.5  ;
estimate 'Female2 vs Male'          type  0  0  0   0    1   -1   ;
```

```
estimate 'Cross2 vs Mid-parent2'      type    0       0       1       0      -0.5    -0.5        ;
estimate 'Cross1 vs Female1'          type    0       1       0      -1       0       0          ;
estimate 'Cross2 vs Female2'          type    0       0       1       0      -1       0          ;
run ;

proc glm data=a ; class entry year rep block type ;
model Noseeds = year rep(year) block(year rep) type entry(type) year*entry(type) ;
lsmeans type / stderr pdiff e=year*entry(type) ;
contrast 'Female1 vs Male'            type    0       0       0       1       0      -1          /e=year*entry(type) ;
contrast 'Cross1 vs Mid-parent1'      type    0       1       0      -0.5     0      -0.5        /e=year*entry(type) ;
contrast 'Female2 vs Male'            type    0       0       0       0       1      -1          /e=year*entry(type) ;
contrast 'Cross2 vs Mid-parent2'      type    0       0       1       0      -0.5    -0.5        /e=year*entry(type) ;
contrast 'Cross1 vs Female1'          type    0       1       0      -1       0       0          /e=year*entry(type) ;
contrast 'Cross2 vs Female2'          type    0       0       1       0      -1       0          /e=year*entry(type) ;
contrast 'Female1 vs Female2'         type    0       0       0       1      -1       0          /e=year*entry(type) ;
estimate 'Female1 vs Male'            type    0       0       0       1       0      -1          ;
estimate 'Cross1 vs Mid-parent1'      type    0       1       0      -0.5     0      -0.5        ;
estimate 'Female2 vs Male'            type    0       0       0       0       1      -1          ;
estimate 'Cross2 vs Mid-parent2'      type    0       0       1       0      -0.5    -0.5        ;
estimate 'Cross1 vs Female1'          type    0       1       0      -1       0       0          ;
estimate 'Cross2 vs Female2'          type    0       0       1       0      -1       0          ;
run ;

proc glm data=a ; class entry year rep block type ;
model SeedLength = year rep(year) block(year rep) type entry(type) year*entry(type) ;
lsmeans type / stderr pdiff e=year*entry(type) ;
contrast 'Female1 vs Male'            type    0       0       0       1       0      -1          /e=year*entry(type) ;
contrast 'Cross1 vs Mid-parent1'      type    0       1       0      -0.5     0      -0.5        /e=year*entry(type) ;
contrast 'Female2 vs Male'            type    0       0       0       0       1      -1          /e=year*entry(type) ;
contrast 'Cross2 vs Mid-parent2'      type    0       0       1       0      -0.5    -0.5        /e=year*entry(type) ;
contrast 'Cross1 vs Female1'          type    0       1       0      -1       0       0          /e=year*entry(type) ;
contrast 'Cross2 vs Female2'          type    0       0       1       0      -1       0          /e=year*entry(type) ;
contrast 'Female1 vs Female2'         type    0       0       0       1      -1       0          /e=year*entry(type) ;
```

```
estimate 'Female1 vs Male'          type  0  0  0  1  0  -1  ;
estimate 'Cross1 vs Mid-parent1'    type  0  1  0  -0.5  0  -0.5  ;
estimate 'Female2 vs Male'          type  0  0  0  0  1  -1  ;
estimate 'Cross2 vs Mid-parent2'    type  0  0  1  0  -0.5  -0.5  ;
estimate 'Cross1 vs Female1'        type  0  1  0  -1  0  0  ;
estimate 'Cross2 vs Female2'        type  0  0  1  0  -1  0  ;
run ;

proc glm data=a ; class entry year rep block type ;
model SeedWidth = year rep(year) block(year rep) type entry(type) year*entry(type) ;
lsmeans type / stderr pdiff e=year*entry(type) ;
contrast 'Female1 vs Male'          type  0  0  0  1  0  -1  /e=year*entry(type) ;
contrast 'Cross1 vs Mid-parent1'    type  0  1  0  -0.5  0  -0.5  /e=year*entry(type) ;
contrast 'Female2 vs Male'          type  0  0  0  0  1  -1  /e=year*entry(type) ;
contrast 'Cross2 vs Mid-parent2'    type  0  0  1  0  -0.5  -0.5  /e=year*entry(type) ;
contrast 'Cross1 vs Female1'        type  0  1  0  -1  0  0  /e=year*entry(type) ;
contrast 'Cross2 vs Female2'        type  0  0  1  0  -1  0  /e=year*entry(type) ;
contrast 'Female1 vs Female2'       type  0  0  0  1  -1  0  /e=year*entry(type) ;
estimate 'Female1 vs Male'          type  0  0  0  1  0  -1  ;
estimate 'Cross1 vs Mid-parent1'    type  0  1  0  -0.5  0  -0.5  ;
estimate 'Female2 vs Male'          type  0  0  0  0  1  -1  ;
estimate 'Cross2 vs Mid-parent2'    type  0  0  1  0  -0.5  -0.5  ;
estimate 'Cross1 vs Female1'        type  0  1  0  -1  0  0  ;
estimate 'Cross2 vs Female2'        type  0  0  1  0  -1  0  ;
run ;

proc glm data=a ; class entry year rep block type ;
model SeedWeight = year rep(year) block(year rep) type entry(type) year*entry(type) ;
lsmeans type / stderr pdiff e=year*entry(type) ;
contrast 'Female1 vs Male'          type  0  0  0  1  0  -1  /e=year*entry(type) ;
contrast 'Cross1 vs Mid-parent1'    type  0  1  0  -0.5  0  -0.5  /e=year*entry(type) ;
contrast 'Female2 vs Male'          type  0  0  0  0  1  -1  /e=year*entry(type) ;
contrast 'Cross2 vs Mid-parent2'    type  0  0  1  0  -0.5  -0.5  /e=year*entry(type) ;
```

```
contrast 'Cross1 vs Female1'        type    0    1    0   -1    0    0    /e=year*entry(type) ;
contrast 'Cross2 vs Female2'        type    0    0    1    0   -1    0    /e=year*entry(type) ;
contrast 'Female1 vs Female2'       type    0    0    0    1   -1    0    /e=year*entry(type) ;
estimate 'Female1 vs Male'          type    0    0    0    1    0   -1    ;
estimate 'Cross1 vs Mid-parent1'    type    0    1    0 -0.5    0 -0.5    ;
estimate 'Female2 vs Male'          type    0    0    0    0    1   -1    ;
estimate 'Cross2 vs Mid-parent2'    type    0    0    1    0 -0.5 -0.5    ;
estimate 'Cross1 vs Female1'        type    0    1    0   -1    0    0    ;
estimate 'Cross2 vs Female2'        type    0    0    1    0   -1    0    ;
run ;

proc glm data=a ; class entry year rep block type ;
model Lscore = year rep(year) block(year rep) type entry(type) year*entry(type) ;
lsmeans type / stderr pdiff e=year*entry(type) ;
contrast 'Female1 vs Male'          type    0    0    0    1    0   -1    /e=year*entry(type) ;
contrast 'Cross1 vs Mid-parent1'    type    0    1    0 -0.5    0 -0.5    /e=year*entry(type) ;
contrast 'Female2 vs Male'          type    0    0    0    0    1   -1    /e=year*entry(type) ;
contrast 'Cross2 vs Mid-parent2'    type    0    0    1    0 -0.5 -0.5    /e=year*entry(type) ;
contrast 'Cross1 vs Female1'        type    0    1    0   -1    0    0    /e=year*entry(type) ;
contrast 'Cross2 vs Female2'        type    0    0    1    0   -1    0    /e=year*entry(type) ;
contrast 'Female1 vs Female2'       type    0    0    0    1   -1    0    /e=year*entry(type) ;
estimate 'Female1 vs Male'          type    0    0    0    1    0   -1    ;
estimate 'Cross1 vs Mid-parent1'    type    0    1    0 -0.5    0 -0.5    ;
estimate 'Female2 vs Male'          type    0    0    0    0    1   -1    ;
estimate 'Cross2 vs Mid-parent2'    type    0    0    1    0 -0.5 -0.5    ;
estimate 'Cross1 vs Female1'        type    0    1    0   -1    0    0    ;
estimate 'Cross2 vs Female2'        type    0    0    1    0   -1    0    ;
run ;

 proc glm data=a ; class entry year rep block type ;
model ascore = year rep(year) block(year rep) type entry(type) year*entry(type) ;
lsmeans type / stderr pdiff e=year*entry(type) ;
contrast 'Female1 vs Male'          type    0    0    0    1    0   -1    /e=year*entry(type) ;
```

```
contrast 'Cross1 vs Mid-parent1'      type    0       1       0      -0.5      0      -0.5      /e=year*entry(type) ;
contrast 'Female2 vs Male'            type    0       0       0       0        1       -1       /e=year*entry(type) ;
contrast 'Cross2 vs Mid-parent2'      type    0       0       1       0      -0.5     -0.5      /e=year*entry(type) ;
contrast 'Cross1 vs Female1'          type    0       1       0      -1        0       0        /e=year*entry(type) ;
contrast 'Cross2 vs Female2'          type    0       0       1       0      -1       0        /e=year*entry(type) ;
contrast 'Female1 vs Female2'         type    0       0       0       1      -1       0        /e=year*entry(type) ;
estimate 'Female1 vs Male'            type    0       0       0       1       0       -1       ;
estimate 'Cross1 vs Mid-parent1'      type    0       1       0      -0.5     0      -0.5      ;
estimate 'Female2 vs Male'            type    0       0       0       0       1       -1       ;
estimate 'Cross2 vs Mid-parent2'      type    0       0       1       0      -0.5     -0.5      ;
estimate 'Cross1 vs Female1'          type    0       1       0      -1        0       0        ;
estimate 'Cross2 vs Female2'          type    0       0       1       0      -1       0        ;
run ;

 proc glm data=a ; class entry year rep block type ;
model bscore = year rep(year) block(year rep) type entry(type) year*entry(type) ;
lsmeans type / stderr pdiff e=year*entry(type) ;
contrast 'Female1 vs Male'            type    0       0       0       1       0       -1       /e=year*entry(type) ;
contrast 'Cross1 vs Mid-parent1'      type    0       1       0      -0.5     0      -0.5      /e=year*entry(type) ;
contrast 'Female2 vs Male'            type    0       0       0       0       1       -1       /e=year*entry(type) ;
contrast 'Cross2 vs Mid-parent2'      type    0       0       1       0      -0.5     -0.5      /e=year*entry(type) ;
contrast 'Cross1 vs Female1'          type    0       1       0      -1        0       0        /e=year*entry(type) ;
contrast 'Cross2 vs Female2'          type    0       0       1       0      -1       0        /e=year*entry(type) ;
contrast 'Female1 vs Female2'         type    0       0       0       1      -1       0        /e=year*entry(type) ;
estimate 'Female1 vs Male'            type    0       0       0       1       0       -1       ;
estimate 'Cross1 vs Mid-parent1'      type    0       1       0      -0.5     0      -0.5      ;
estimate 'Female2 vs Male'            type    0       0       0       0       1       -1       ;
estimate 'Cross2 vs Mid-parent2'      type    0       0       1       0      -0.5     -0.5      ;
estimate 'Cross1 vs Female1'          type    0       1       0      -1        0       0        ;
estimate 'Cross2 vs Female2'          type    0       0       1       0      -1       0        ;
run ;
```