Application of molecular and bioinformatics approaches for sensory and quality attributes improvement in peanuts (*Arachis hypogaea* L.)

By

Tao Jiang

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 14, 2019

Keywords: peanut, sensory and quality attributes, MAS, association mapping, GWAS, RNA-seq

Approved by

Charles Chen, Chair, Professor of Crop, Soil and Environmental Sciences
Jacek Wower, Professor of Animal and Dairy Sciences
Scott McElroy, Professor of Crop Soil and Environmental Sciences
Phat Dang, Research Biochemist, USDA-ARS, National Peanut Research Laboratory
Xu Wang, Assistant Professor of Animal Health

Abstract

As more consumers are paying attention to food quality and nutrients, obtaining desirable flavor and quality attributes become an important breeding objective for peanut. In this study, 120 accessions from the US Mini-core collection are utilized to investigate the population structure and genome-wide association studies (GWAS) are employed to identify its impacts on the market-trait association. In order to analyze the flavor and quality attributes (including sensory attributes, sugars, fatty acids, and tocopherols), this study first used the Affymetrix version 2.0 SNP array to identify a total of 17,232 high-quality single nucleotide polymorphisms (SNPs) from the U.S. peanut Mini-core Collection. 310, 474, 569 and 230 QTLs were identified related to sensory attributes, sugars, fatty acids, and tocopherols, respectively, including 14, 87, 105 and 73 significant QTLs for each profile. Among the significant QTLs, there were 4, 55, 86, and 49 QTLs with PVE higher than 15% for each profile, respectively (Table 3), indicating the high heritability of these traits. Consequently, genes were screened within 1 Mb of the significant QTLs and a total of 136 candidate genes were identified functionally associated with corresponding traits, including 13 genes associated with sensory attributes, 65 genes with sugars, 35 genes with fatty acids and 23 genes with tocopherols, respectively. Among these candidate genes, most of them are known to be involved in the biosynthesis and metabolism process of carbohydrates, fatty acids, and polyphenols. Furthermore, these results were confirmed by association mapping using SSRs and next-generation RNA-sequencing (RNA-Seq) two approaches. These findings provided a promising insight into the complicated genetic architecture of sensory and quality attributes in peanut and would promote the marker-assisted selection (MAS) in peanut breeding with seed flavor and nutrition quality.

Acknowledgments

I would like to express my highly respect and thanks to my major advisors, Dr. Charles Chen, for both the academic guidance, encouragement and support during the past four years. Without his patience, this thesis would never be accomplished. I am deeply grateful for the time, help, and effort from my committee members: Dr. Jacek Wower, Dr. Scott McElroy, Dr. Phat Dang, and Dr. Xu Wang.

I gratefully appreciate the whole personnel from the E.V. Smith Research Center, Auburn University, Headland, AL, Wiregrass Research and Extension Center, Auburn University, Headland, AL, and from the USDA-ARS National Peanut Research Lab at Dawson, GA for the great helps with all experimental affairs related to my research.

I am also grateful to my fellow graduate students as well as visiting scholars in Dr. Chen' lab: Dr. Yueyi Tang, Jing Li, Xu Wang, Dr. Shuli Zhao, Dr. Shuzhen Zhao, Lauren Lang, Dr. Hui Zhang, Yan Yu, Oluwamosope Adeniji, Boipelo Nthupisang, Li Li, John Bagwell, Dr. Xinlei Yang, and Dr. Juan Liu, for their assistance with my laboratory work and personal life.

Finally, I would like to thank my beloved wife, Zhao Peng, my parents, Mr. Zaihao Jiang and Mrs. Weiduo Zhong, and my sisters, Miaomiao Jiang and Yuanyuan Jiang for all their supports, encouragement and companionship. Their guidance and encouragement inspire me in my whole life.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AFLP	amplified fragment length polymorphism
ANOVA	analysis of variance
CAF	correlated allele frequencies
DSA	descriptive sensory analysis
DEG	differential gene expression
DR	dark roast
DPPH	2, 2-diphenyl-1-picrylhydrazyl
EDTA	ethylenediaminetetraacetic acid
EMMAX	efficient mixed-model association eXpedited
EST	expressed sequence tag
FF	fruity fermented
F*st*	fixation index
FPKM	fragments per kilobase of transcript per million mapped reads
GLM	general linear model
GWAS	genome-wide association study
HR	hypersensitive response
InDel	insertion and deletion
ISSR	inter simple sequence repeat
IV	iodine value
LCS	percentage of low level of calcium soaps of palm oil fatty acid
LD	linkage disequilibrium

MAS	marker-assisted selection
MCMC	Markov Chain Monte Carlo
MLM	mixed linear model
mM	millimolar
ng	nano gram
NGS	next generation sequencing
OAV	odor activity value
O/L ratio	ratio of oleic acid (18:1) to linoleic acid (18:2)
ORF	open reading frame
PC	plastic chemical
PCA	principal component analysis
PCR	polymerase chain reaction
PIC	polymorphic information content
P/S ratio	ratio of palmitic acid(16:0) to stearic acid (18:0)
QTL	quantitative trait loci
RAPD	random amplified polymorphic DNA
RB	raw beany
RCB	randomized complete block
RFLP	restriction fragment length polymorphism
RIL	recombinant inbred line
RNA-Seq	RNA-sequencing
ROC	percentage of raw oil content (wet weight) in total weight
RP	roast peanutty

SA	sweet aromatic
SAS	Statistical Analysis Software
Sat (%)	percentage of saturated fatty acid of total fatty acid
SNP	single nucleotide polymorphism
SPAGeDi	Spatial Pattern Analysis of Genetic Diversity
SSR	simple sequence repeat
SVS	SNP & Variation Suite
TSWV	tomato spotted wilt virus
UPGMA	unweighted pair group method with arithmetic average
USDA	United States Department of Agriculture
wt/vol	weight per volume
WHS	woody hulls skins
ΔK	second order rate of change of the likelihood

Chapter One

Literature Review

1.1 Introduction

As one species in the legume family, which bears small and bright yellow flowers, and a subterranean pod, the peanut (*Arachis hypogaea* L., 2n = 4x = 40), produces edible seeds. It can be classified into nine taxonomic sections based on its morphology, geographic distribution, and cross-compatibility (Bertioli et al., 2011; Upadhyaya et al., 2011; Pandey et al.,2012). Cultivation can produce a spontaneous chromosome duplication hybrid, such as A. *hypogaea* is cross produced by the wild diploid species A. *duranensis* (A-genome) and A. *ipaensis* (B-genome) (Seijo et al., 2004). Peanut is genetically diploid, even if it is a tetraploid (AABB) (Stalker et al., 1991). The DNA of domesticated peanut is equally attributed to the A- and B-genomes (Singh et al., 1996).

As the worldly third-largest oil-seed crop, the production of peanut can reach to 41.3 million tons and 1676 kg/ha. Nevertheless, the production of Asia (2217 kg/ha) and Africa (929 kg/ha) is lower than the Americas (3632 kg/ha) (FAOSTAT, 2014). In 2018, approximately 5.46 million metric tons of peanuts were produced in the U.S.A. (USDA, 2019). Not only limited to the peanut oil, cake, and meal, the U.S. peanuts industry also produces toasted peanut, peanut butter, savory snacks, peanut oil and peanut flour (Kaya et al., 2009). Half of the peanut production market is peanut butter, and the other half is snack nuts and confectionery products (American Peanut Council, 2017). The unique and pleasant flavor of peanuts during roasting is the reason for its popularity (Sanders et al.,1997). Many factors can influence peanut flavor, including genetic, environmental, handling, processing, and storage conditions(Sanders et al.,1995). The interactions among these factors can affect the chemical composition of peanuts and then affect its flavor quality. For example, over 200 aroma-active compounds await to be identified to have an influence

over the roasted peanuts' flavor (Schirack et al., 2006). Hence, it is crucial to explore how the aroma-active compounds affect the peanuts' flavors, especially the specific compounds which are responsible for the roasted peanut flavor. Hence, how to obtain desirable flavor and quality attributes becomes an important breeding objective for peanut.

However, although a massive effect through traditional breeding methods has been made to improve crop productivity, the outcomes are still unsatisfactory. The increasing demand for peanut requires higher productivity; thus, conventional breeding is no longer adequate to address the complicatedgenetic behavior of the majority of aspired traits. Even though plant and crop breeding researchers are still dedicating to find out the best methods of molecular plant breeding (Khedikar et al., 2010; Holbrook et al.,2011; Chu et al., 2011; Pandey et al., 2012; Fountain et al., 2015; Song et al., 2016; Janila et al., 2016; Shasidhar et al., 2017), there require more perspective from biology basic scientific research.

The present study aims to provide an overview of the significant milestone developments in molecular plant breeding history, the major principles that influence the current practice, and the factors that influence the adoption process in peanut improvement programs. This study made a significant contribution to discovering the application of genes and their function on molecular plant breeding.

1.2 The sensory and quality attributes of peanut

Different varieties of macro- and micro- nutrients constitute the human diet, with multiple types of vitamins, minerals, antioxidants, and other beneficial phytochemicals. All the above are indispensable to maintain a healthy body. The macronutrients are sources to proffer all sorts of proteins, carbohydrates, and fats; in the meantime, the food industry is responsible for supplying these as primary products or as ingredients of a wide variety of foods (Gunstone, 2011). Previous

studies have been intensely focused on the chemical components of cultivated peanut; however, it remains unclear and waits for further elucidation as to how the chemicals or processing steps affect the ultimate peanut flavor and nutrition. Roasting peanuts can produce a unique flavor, which is one of the significant factors influencing consumers' choice and acceptance. Even though it has been considered for 40 years that pyrazines are the source of roasted peanut flavor, the reliability and validity of this fact need more information to substantiate, which poses a challenge to the investigation in peanut flavor quality. Moreover, pyrazines found in the volatile composition of roasted peanut does not necessarily indicate that they play a significant role in peanut flavor. When the concentrations are below the sensory threshold, there is a possibility that compounds in a portion of food do not have any influence in deciding the flavor (McGorrin, 2002). Besides, prior studies have reported that pyrazines producing nutty/roasted notes do not explicitly imply they are the basis for roasted peanut flavor. Sometimes the interactions with the matrix and other chemical compounds in the food may affect the relationship between the individual aroma compounds and the corresponding flavor. (Drake and Civille, 2003). The techniques used to isolate and separate the flavor components may lead to the formation of several artifacts and loss of volatiles. There are no defined references, language, or system developed or carried out to identify the role of pyrazine compounds in roasted peanuts' flavors, except the Odor activity value (OAV), which is the ratio of concentration to the sensory odor threshold (Drake, 2007). Therefore, further measures and analysis require to be taken to prove and clarify the importance and balance of these compounds for the overall flavor of roasted peanuts, since there is no report about OAVs isolated from the flavor compounds in early studies.

1.2.1 Flavors profiles

Widely used lexicon terms to analyze the flavor characteristics is descriptive sensory analysis (DSA). DSA was developed by a group of experienced panelists and has often been used to describe the various flavor characteristics. When it comes to the completion phase, a well-defined and non-random list of flavor terms will be used to describe the newfound sensory characteristics from a particular food (Drake and Civille, 2003). The other advantage of DSA is to have a panel trained in using the lexicon descriptors, which function to evaluate both quantitative and qualitative components of a product (Meilgaard et al.,1991). The participation of a professionally trained panel can contribute to the efficiency of final results, such as the compelling correlation of DSA with instrumental and/or consumer data that can be reached(Drake and Civille, 2003).

In general, there are three steps determining the relationships between sensory and instrumental analysis of foods: 1) selection of flavors of interest using DSA; 2) instrumental analysis of volatile extracts; and 3) confirmation of aroma -active compounds via quantitation, threshold testing, and DSA of model systems (Drake et al.,2006). Schirack et al. (2006) used this approach and found that phenylacetaldehyde, guaiacol, and 2,6 - dimethyl pyrazine were responsible for the stale/floral and ashy off - flavoring high-temperature microwave - blanched peanuts. By following a similar tactic, Didzbalis et al. in 2004 showed that fruit - like esters such as ethyl 2 - methyl propanoate, ethyl 2 - methyl butanoate, and ethyl 3 - methyl butanoate as well as short-chain in organic acids (butanoic, 3 - methlbutanoic, and hexanoic) were significantly related to the fruity/fermented off-flavor developed in immature peanuts cured under high temperatures.

A well-accepted explanation is that when under heat, foods with roast flavor will have the Maillard reaction, in which the polymerization of free amino acids with sugar can change the flavor (Nursten et al., 2005). These compounds have been detected in peanuts (Chiou, et al., 1993; Pattee et al., 2001; Liu et al., 2011). Besides the outside environment and chemicals inside the food, data

showed that the processing is crucial in producing flavor. For instance, maceration significantly increased roasted peanut and dark roast aromas, and reduce sweet, raw bean aromas, and sweetness. Baking reduces roasted peanut and dark roast, and increases raw bean aromas comparing with frying, and also impacts color development (Lykomitros et al.,2016). Therefore, both raw materials and processing methods can change the flavor and nutrient of peanuts. Based on the previous research, the peanut sensory attributes have high inheritability in peanut, such as attribute sweet, bitter, and sweet aromatic, while quality attributes have even higher (Pattee et al., 1995, 2000). It is reported that the sweet attribute has very high broad-sense heritability with H = 0.28, that is, 28% of the total variability in sweet is due to genetic causes, while heritability of roasted peanutty attribute is only 0.06 and heritability of bitter attributes is 0.06 (Pattee et al., 2001). Except for the sensory attributes, the seed nutrient quality attributes, such as sugars, fatty acids, and tocopherols, are also common study subjects due to their close relation with human health. As reported, genotypes accounted for 38-78% of the total variation for the known Carbohydrates, suggesting that broad-sense heritability for these components is high. (Pattee, 2000). While the levels of tocopherols have been reported to be environmentally influenced, although there may be some genetic forces in effect. (Baydar et al., 2005; Jonnala et al., 2006). Besides, Zhang et al. (2004) showed that the main genetic effects for oleic acid content were more important than genotype × environment interaction effects, with total narrow-sense heritability as high as 77.5% in Brassica napus L.

1.2.2 Sugars profile

The flavor of the roasted peanut (*Arachis hypogaea* L.) seed is an essential factor in influencing consumer acceptance. As an antecedent, carbohydrates can facilitate the process in which the compounds impart the roasted peanut characteristic (Neta et al., 2010; Davis et al., 2015).

Although the function of the peanut carbohydrates has been identified and quantitated (Pattee et al., 2000; Neto et al., 2010; Shiriki et al., 2015), we are still unclear about their specific role in roasted peanut flavor generation. For example, after looked into peanut monosaccharide, oligosaccharide, and polysaccharide fractions as the origin of intermediates for flavor molecules, Cha et al. in 1993 found that heating will result in many identified sensory. It was found that the individual components of the peanut carbohydrate fraction would change during the phase of maturation and curing (Ross and Mixon, 1989; Vercellotti et al., 1995). They also change across seed size and over storage time (Pattee et al., 1981), decrease higher soil temperatures (McMeans et al., 1990), and vary among a limited number of genotypes (Basha, 1992). What kind of genotypic relationship caused by such variations remains yet unknown. The sensory attributes of roasted peanut are heritable and subjected to environmental conditions (Pattee and Giesbrecht, 1990; Pattee et al., 1997, 1998; Isleib et al., 1995). The following research focused on exploring genotypic carbohydrate variation needs to address the role of carbohydrates in roasted peanut quality development.

In summary, there is a wide range of genotypic variation in carbohydrate components across different peanut genotypes. Precedent work has shown similarly high genotypic variation for the sweetness attribute in roasted peanuts (Pattee et al., 1998). A higher degree of sweetness usually results in superior flavor profiles, compared with the lower bitter and higher roasted peanut attribute intensities. As the sensory evaluation is a costly process, it would be rewarding to establish the relationships between sensory attributes and chemical composition in peanut breeding for the purpose of keeping the flavor quality in future peanut varieties. The indirect selection for flavor may be more efficient than direct selection, as the former is based upon a simple chemical assay of carbohydrate components, and the latter relies upon sensory panel data.

1.2.3 Fatty acids profile

The palmitic acid (C16:0), oleic acid (C18:1), and linoleic acid (C18:2) are the primary fatty acid in peanut oil. However, other acids, such as stearic acid (C18:0), arachidic acid (C20:0), eicosenoic acid (C29:1), behenic acid (C22:0), and lignoceric acid (C24:0) are in minimum quantities. Linoleic fatty acids (C18:3) present in trace amounts (Casini et al., 2003). The Codex Alimentarius suggests that in peanut oil, the arachidic and higher fatty acid content should not exceed 48 g/kg. Other fatty acids, such as the C20:0 and C22:0 in olive oil, should be following adulteration with peanut oil (Young, 1996). Peanut oil is generally composed of 80% unsaturated fatty acids and 20% saturated ones (Cobb and Johnson, 1973). The oil in mature peanuts contains 96% triacylglycerol, with the main fatty acids being palmitic, oleic, and linoleic (Ahmed and Young, 1982; Sanders, 1980). In peanut oil, the percentage of long-chain fatty acids should be approximate to or smaller than 2%, and the percentage of FFAs should stay between 0.02 and 0.6% (Guthrie et al., 1949). The overview of the fatty acid profile of peanut oil is presented in Table 3. Peanut oil is an ideal option for cooking and frying due to its excellent oxidative stability (O'Brien, 2004). Soybean oil may be a better choice when frying, but it could generate some flavor defects when in long-term use (Young, 1996). The oleate/linoleate ratio (O/L) and iodine value (IV) are important, especially in terms of governing product shelf life. Low IV and high-O/L ratio indicates enhanced shelf life and decreased rancidity (Andersen and Gorbet, 2002). Different genotype, seed maturity, climatic condition, growth location will generate different fatty acid compositions of peanut oil (Hashim et al.,1993). During the seed growing period, the lower temperatures would result in more unsaturation (Casini et al., 2003). Generally, when the oleic acid increases, the linoleic would decrease as the seed arrives at maturity. When the seed progresses from intermediate to mature stage, palmitic and linoleic acids decrease while oleic acid increases (Hinds, 1995). Some other

studies showed that the fatty acid profile is stable across years (Andersen and Gorbet, 2002). The soil type can be an influential factor, and prior studies found that seeds would contain more stearic acid and saturated fatty acids when grown on the volcanic clay loam, but would contain less linoleic and total unsaturated acids if grown on volcanic sandy loam (Hinds, 1995). Worthington et al. (1972) also found that genotypes can also influence the fatty acid composition of peanuts. Generally, the three major fatty acids (Palmitic, oleic, and linoleic acid) were more pliant to outside factors than the minor fatty acids. In particular, Isleib et al. (2006) showed that the oleic-to-linoleic acid ratio (O/L ratio) was over ten times higher for HO peanuts comparing with normal peanuts. This change in fatty-acid profile was due to the incorporation of two key recessive genes into breeding lines, yielding new high HO varieties (Knauft et al., 1993).

1.2.4 Tocopherols profile

Tocopherol is one of the natural antioxidants. Generally, peanuts oil contains four types of tocopherols (Carr´ın and Carelli, 2010). Noticeably, tocopherols, and tocotrienols can be removed by 10–20% through refining caustic and alkali chemical, and 30–60% can be removed through deodorization or steam distillation (O'Brien, 2004). Therefore, α- and γ -tocopherol is mostly found in refined peanut oil. Casini et al.(2005) investigated four harvest periods and found that the tocopherol levels between 199 and 815.6 ppm to crude type runner peanut oil from Argentina. They, therefore, concluded that tocopherol content would increase under the condition of high precipitations and low soil temperatures. Similarly, Hashim et al. (1993) also discovered that there are significant differences between maturity stages of runner- and virginia-type peanut cultivars within the tocopherol profile. Prior data, which is from an oil composition factors study, demonstrated that the tocopherol content varies consistently across peanuts from various origins (Sanders et al., 1992). Compared to the peanuts grown in China or Argentina, peanuts grown in

the U.S. can produce higher tocopherol content. The highest levels of a whole seed basis were 250 ppm, and the lowest levels were about 100 ppm. Misuna et al.,(2008) found that the antioxidant activity of eight commercial peanut cultivars by 2, 2-diphenyl-1-picrylhydrazyl (DPPH) assay and suggested that peanut oil has an antioxidant activity of 0.17% (W/V).

Hence, with these high hereditary characters above, it could be more efficiently and effectively to conduct the peanut breeding based on sensory and quality attributes by obtaining genetic molecular markers associated to corresponding traits and utilizing them in Marker-assisted selection, comparing to the long-term and high-cost traditional breeding strategy (Collard et al., 2005, 2007; Xu et al., 2008).

1.3 Marker-assisted selection (MAS)

Mapping out and tagging essential crop genes can be significantly facilitated along with a range of crop plants molecular markers (Ibitoye, et al.,2010). The importance of marker-assisted selection (MAS) should be attenuated as it can improve the efficiency of plant breeding through a precise transfer of the genomic regions of interest and facilitate the recovery of the recurrent parent genome. The success of MAS is under the influence of a few critical factors, for example, the amount of target genes to be transferred, the range between the target gene and the flanking markers, the amount of genotypes chosen in each breeding propagation, the characteristics of germplasm and the technical options available at the marker level (Kumar, 1999; Collard et al.,2005).

As reported in some temperate cereal crops (Varshney et al.,2006) and legume crops (Varshney et al.,2010a), molecular genetic breeding exceeds the conventional breeding on handling traits, which is challenging to administer during conventional phenotypic selection. In addition to handling trait complexity, molecular breeding can more accurately introgress multiple recessive alleles with less

time spent along with pyramiding of several monogenic traits or QTLs for a single trait (Xu and Crouch, 2008; Varshney et al.,2009). However, having the large amount of molecular markers, genetic maps, and suitable phenotyping platforms is the prerequisite to apply the molecular breeding approaches on germplasm. Furthermore, with all the efforts united from the global peanut academicians, a significant progress has been made in the area of *Arachis* genomics and some efforts have been spent on QTL mapping and molecular breeding for resistance to biotic/abiotic stresses for peanut improvement in the past five years (Varshney et al.,2010). Among all genomic resources, molecular markers can be directly applied to germplasm characterization, trait mapping, and molecular genetic breeding. Many advanced marker systems, such as restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNAs (RAPDs), amplified fragment length polymorphisms (AFLPs), diversity arrays technology (DArT) markers, sequence-tagged sites (STSs), and expressed sequence tags (ESTs) (Kesawat et al.,2009; Gupta et al.,2010; Ibitoye et al., 2010), have been developed during the past three decades.

Nevertheless, simple sequence repeats (SSRs) or microsatellites and single nucleotide polymorphism (SNP) markers are regularly advocated for plant genetics and breeding strategies. SSR markers are multi-allelic, co-dominant, and operable, while the SNP markers are highly vulnerable to high-throughput genotyping methods. However, developing and employing SNP markers is still not routinely used as a technique in crop species and especially not in low-tech laboratories. In peanut, some tryouts have been made such as using RFLP, RAPD and AFLP markers to do diversity analysis (Subramaniyan et al.,2000; Herselman, 2003; Bravo et al.,2006) and genetic mapping (Herselman et al.,2004; Garcia et al.,2005; Leal-Bertioli et al.,2009). However, only a small amount of SSR markers were available until 2005. The low diversity

identified by SSR markers in the cultivated gene pool indicates that in future studies, we need a large-scale SSR marker development.

Besides, transcriptomic resources, for example, expressing sequence tags (ESTs), can help us understand and explore genome dynamics, the development of gene-based markers and maps, the transcript profiling for identifying the candidate genes involved in the expression of traits of interest, and the identification of transcriptional changes during plant immunity responses. Currently, the NCBI has a total of 252,832 *Arachis* ESTs are available in the public domain (Guo et al., 2015). The University of Georgia has developed almost 1 million reads representing >350 Mb of transcript sequences from 17 tetraploid genotypes by using 454-titanium sequencing technology.

With advancements in genetic testing methods, more simple breeder-friendly markers are coming in the spotlight. These technological improvements will bring screening for allele composition closer to breeding programs (Ibitoye et al.,2010). In addition, to deeply understand the potential function of markers in genetic improvement of crops, improvements in genomics of model species such as rice and Arabidopsis should be integrated with DNA marker technology.

1.4 Molecular approaches for peanut breeding

1.4.1 Association mapping using SSRs

A quantitative trait locus (QTL) is a section of DNA which correlates with variation in a quantitative phenotype. The QTL links typically to, or contains, the genes that control the phenotype. Using QTL mapping can help identify markers for marker-assisted selection (MAS) and therefore enhance the efficiency of selection in plant breeding (Li et al., 2011). There are two approaches to genetic mapping. One is traditional bi-parental mapping, which draws the population from crosses of two individual genotypes. However, the restricted recombination

events restrict genetic resolutions. Moreover, it would cost more time and effort for the bi-parental mapping approach to create large enough populations. Besides, it requires more steps to pinpoint the location of the QTL or clone the gene. The other approach is called association mapping, which uses the recombination events from many lineages, and exploits relationships between markers and traits in much broader germplasm. The first step of association mapping is to identify the markers with the unequal distribution of alleles among the individuals of a trait. Then the researcher can identify the markers that have the association with a particular trait of interest (Schafer and Hawkins, 1998). Comparing with the traditional mapping, the association mapping takes less time and money and increases the precision of QTL location. Moreover, because of the association analysis using broad germplasm with a diverse genetic background, it is more efficient to identify the association between markers and traits (Zeng et al., 2009). The linkage disequilibrium (LD), which association mapping relies on, is the physical linkage between markers and trait loci.

However, different from F2 or RIL (recombinant inbred line), the observed LD in a natural or breeding population also could be produced by non-linkage factors like genetic drift, selection and population admixture (Jannink and Walsh, 2002; Flint-Garcia et al., 2003; Mackay and Powell, 2007). Thus, LD could be observed between unlinked loci, which could cause false-positive associations. The number of false associations can be reduced by comprising the population structure and kinship relatedness among individuals (Stich et al., 2008). In addition, this problem also can be partially solved by using random mating populations. The LD between QTLs and unlinked markers could decrease to some extent when the recombination rate increases (Zeng et al., 2009).

Association analysis was first applied in humans to identify loci causing certain diseases (Risch and Merikangas, 1996). In plants, this method was first used to identify associations with flowering time in maize (Thornsberry et al., 2001). Currently, association mapping has been implemented in above ten plant varieties, such as glutinous phenotype and yield in rice (Olsen and Purugganan, 2002; Agrama et al., 2007), heading date and flowing time in perennial ryegrass (Skøt et al., 2005, 2007), late blight resistance in potato (Malosetti et al., 2007), salt tolerance in wild barley (Ivandic et al., 2003) and fiber traits in cotton (Kantartzi and Stewart, 2008). In peanut related studies, Wang et al. (2011) were the first ones to make an attempt at association mapping. Following his first tryout, Yan et al. (2013) applied 64 SSR markers in 136 accessions and found four markers associated with quality traits based on 3-years phenotyping data. More lately, a genome-wide association study (GWAS) in peanut has been reported, in which 300 genotypes were tested for 36 traits, including biotic and abiotic defense, seed quality, and yield. (Pandey et al., 2014).

Based on previous studies, this study applied 81 SSR markers and two functional SNP markers from FAD2 on 94 *Arachis hypogaea* Mini-core collection germplasm accessions for association mapping of seed quality traits in peanut.

1.4.2 Genome-wide association studies (GWAS)

Most natural variants in crop plants are generated from spontaneous mutations in wild progenitor plants under both natural and human selection. (Huang et al.,2014). By combining the colossal germplasm resources with genetic tools such as genome sequences, genetic populations, transformation techniques and genome-wide association studies (GWAS), crop researchers have the capability to extensively and efficiently make use of natural variation and associated phenotypic variation with the underlying genotypic variants (Godfray et al.,2010; Tester et al.,2010). With the rapid development of sequencing technologies, such as diversity arrays

technology and computational methods, GWAS is now becoming a powerful tool for detecting natural variation underlying complex traits in crops (Rafalski, 2010). Different form GWAS in humans, which usually adopt a case-control design (Int. HapMap Consort, 2005& 2007), GWAS employed in crops, usually relies on a permanent resource and can be used to generate specific mapping populations for specific traits or QTLs in crops (Atwell et al., 2010).

Heretofore, there have been many successfully cases of GWAS being carried out in many crops, including maize, rice, sorghum, and foxtail millet (Huang et al.,2010; Huang et al., 2012; Jia et al., 2013; Kump et al.,2011; Li et al., 2013; Morris et al.,2012; Zhao et al., 2011). Based on the size of developed and published resources, rice and maize are the two primary models for crop GWAS. In peanut, with sequence completion of the two progenitors of cultivated peanut, *Arachis duranensis* and *Arachis ipaensis* and Mini-core large amounts of the sequence variation in peanut germplasm captured by United States Department of Agriculture (USDA) (Fronicke et al.,2011). The SNPs identified in the data will be analyzed for the linkage disequilibrium evident in the diversity panels, and therefore, provide the foundation for GWAS studies (especially for the phenotypes already measured) and efficient QTL mapping as well as the generation of a peanut haplotype map in conjunction with the reference sequence.

A matter requiring attention in crop GWAS analysis is the need to account for the massive population structure, including weighing the trade-offs of increased false-negative and decreased false-positive rates from accounting for that structure (Korte et al.,2013; Platt et al.,2010).

There are some powerful tools for testing associations between single-nucleotide polymorphisms and multiple correlated phenotypes in genome-wide association studies, including the Efficient Mixed-Model Association eXpedited (EMMAX) program, the compressed mixed linear model method and the interactive Web-based application, GWAPP (Kang et al.,2010, Lipka et al.,2012;

Lippert et al.,2011; Seren et al.,2012). However, of the variety of genes in the middle of one GWAS locus, only one may function to the QTL, which is known as the causal gene. Hence, the following analyses of GWAS loci and supplementary experiments might be required to pinpoint the causal genes. Gene annotation, expression profiles, and variant catalogs can be thoroughly analyzed in the post-GWAS stage.

1.4.3 Next-Generation RNA-Sequencing (RNA-Seq)

RNA-Seq is a lately developed strategy for analyzing transcriptome profiles that applies deep-sequencing methods, rendering a much more accurate measurement of levels of transcripts and their isoforms. (Wang et al.,2009). It is rapidly overwhelming microarrays as the technology of choice for whole-transcriptome studies (Van Verk et al.,2013). For example, it is a critical and efficient method to catalog expressed genes, validate gene predictions, and profile gene expression in cultivated peanut tissues across multiple developmental stages and upon challenge with multiple stresses. RNA-Seq is a more extensive and precise measurement of gene expression than microarray experiments. Similar study is being done in soybean, and an example of the study can be obtained at http://www.soybase.org/soyseq/ (Froenicke et al.,2011). The advent of RNA-sequencing (RNA-Seq), a rapid technique for genome-wide gene expression analysis, provides powerful alternatives to facilitate the production of different genotypes in a more efficient manner (Lu et al., 2014). Recently, gene structures and expression profiles in many crops, including wheat, corn, soybean, peanut, in response to stress conditions were determined with RNA-Seq technology (Brasileiro et al., 2015; Chen et al., 2013; Li et al., 2014; Lang et al., 2019; Petre et al., 2012; Ruan et al., 2018; Zhao et al., 2018).

Generally, we converted mRNA into cDNA and then form an RNA-Seq library in an RNA-Seq experiment. We can achieve relative abundant and accurate measured transcript and splice variants

by sequencing the millions of DNA fragments in the library. VanVerk et al.,(2013) provided detailed instructions on how to systematically implement these tools at each step of a standard plant RNASeq workflow (showed as followed) for plants with a reference genome.

The use of RNA-Seq for genome-wide transcriptome studies has increased drastically in the past years. If transcripts can be sequenced without fractioning the cDNA or amplifying PCR, this tool would primarily reduce computing time and improve the correct assignment of sequencing reads. This tool can bring possibilities for whole-genome expression profiling at an unprecedented level of detail (Majewski et al.,2011).

The above four sections emphasize that despite recent advances and successful examples in molecular crop breeding, one of the current notable challenges remains to identify those gene combinations that lead to significant peanut improvement. It also suggests that the most effective approach to accelerate such efforts is to integrate better the different research disciplines and activities that form core components of molecular crop breeding.

1.5 Overall goal and objectives

The goal of this research was to screen the U.S. peanut Mini-core collection for sensory attributes and detect the QTL responsible for crop improvement. In particular, the objectives of this research were to

Objective 1: Population structure and molecular marker-trait association mapping of sensory attributes using SSRs in peanut.

1. Examine genetic diversity and population structure in the peanut Mini-core collection by SSR markers;
2. Conduct association mapping of SSR markers to sensory attributes in peanut;
3. Provide genetic materials for breeding programs.

Objective 2: Genome-wide association study of sensory attributes using SNPs in peanut.

1. Examine genetic diversity and population structure by SNP markers
2. Locate SNP markers associated with the traits for MAS breeding
3. Narrow down genes associated with SNPs involved in corresponding traits and study the pathway related to the responses.

Objective 3: Identify and functional validate genes involved in high oleic acid using high-throughput mRNA sequencing in cultivated peanut.

1. Identify genes regulated in different growth stages.
2. Identify genes related to high oleic acid trait by comparing the four genotypes.
3. Reveal and analyze the post-transcriptional regulation pathway related to high oleic acid.

1.6 References

Agrama, H. A., Eizenga, G. C., & Yan, W. (2007). Association mapping of yield and its components in rice cultivars. Molecular Breeding, 19(4), 341–356. https://doi.org/10.1007/s11032-006-9066-6

American Peanut Council. (2017). The Peanut Butter Portraits. https://peanutbutterportraits.wordpress.com/the-industry/. Accessed Date June 1, 2017.

Andersen, P. C., & Gorbet, D. W. (2002). Influence of year and planting date on fatty acid chemistry of high oleic acid and normal peanut genotypes. Journal of Agricultural and Food Chemistry, 50(5), 1298–1305. https://doi.org/10.1021/jf0113171

Andersen, P. C., Hill, K., Gorbet, D. W., & Brodbeck, B. V. (1998). Fatty Acid and Amino Acid Profiles of Selected Peanut Cultivars and Breeding Lines. Journal of Food Composition and Analysis, 11(2), 100–111. https://doi.org/10.1006/jfca.1998.0565

Araus, J. L., & Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. *Trends in Plant Science*, *19*(1), 52–61. https://doi.org/10.1016/j.tplants.2013.09.008

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., … Nordborg, M. (2010). Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. Nature, 465(7298), 627–631. https://doi.org/10.1038/nature08800

Bao, Z. S., Chen, H. M., Yang, M. Y., Zhang, C. B., Yu, K., Ye, W. L., … Jiang, T. (2014). RNA-seq of 272 gliomas revealed a novel, recurrent PTPRZ1-MET fusion transcript in secondary glioblastomas. Genome Research, 24(11), 1765–1773. https://doi.org/10.1101/gr.165126.113

Basha, S. M. (1992). Effect of Location and Season on Peanut Seed Protein and Polypeptide Composition. Journal of Agricultural and Food Chemistry, 40(10), 1784–1788. https://doi.org/10.1021/jf00022a011

Basha, S. M. (1992). Soluble Sugar Composition of Peanut Seed. Journal of Agricultural and Food Chemistry, 40(5), 780–783. https://doi.org/10.1021/jf00017a015

Baydar, H., & Erbaş, S. (2005). Influence of seed development and seed position on oil, fatty acids and total tocopherol contents in sunflower (Helianthus annuus L.). Turkish Journal of Agriculture and Forestry, 29(3), 179–186. https://doi.org/10.3906/tar-0406-3

Bertioli, D. J., Seijo, G., Freitas, F. O., Valls, J. F. M., Leal-Bertioli, S. C. M., & Moretzsohn, M. C. (2011). An overview of peanut and its wild relatives. Plant Genetic Resources, 9(1), 134–149.

Bolotin, A., Quinquis, B., Sorokin, A., & Dusko Ehrlich, S. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology, 151(8), 2551–2561. https://doi.org/10.1099/mic.0.28048-0

Bravo, J. P., Hoshino, A. A., Angelici, C. M. L. C. D., Lopes, C. R., & Gimenes, M. A. (2006). Transferability and use of microsatellite markers for the genetic analysis of the germplasm of some *Arachis* section species of the genus *Arachis*. Genetics and Molecular Biology, 29(3), 516–524. https://doi.org/10.1590/S1415-47572006000300021

Cantonwine, E. G., Culbreath, A. K., Stevenson, K. L., Kemerait, R. C., Brenneman, T. B., Smith, N. B., & Mullinix, B. G. (2006). Integrated disease management of leaf spot and spotted wilt of peanut. Plant Disease, 90(4), 493–500. https://doi.org/10.1094/PD-90-0493

Carrín, M. E., & Carelli, A. A. (2010). Peanut oil: Compositional data. European Journal of Lipid Science and Technology, 112(7), 697–707. https://doi.org/10.1002/ejlt.200900176

Carunchia Whetstine, M. E., & Drake, M. (2007). The flavor and flavor stability of skim and whole milk powders. ACS Symposium Series, 971, 217–251. https://doi.org/10.1021/bk-2007-0971.ch013

Casini, C., Dardanelli, J. L., Martínez, M. J., Balzarini, M., Borgogno, C. S., & Nassetta, M. (2003). Oil quality and sugar content of peanuts (*Arachis hypogaea*) grown in Argentina: Their relationship with climatic variables and seed yield. Journal of Agricultural and Food Chemistry, 51(21), 6309–6313. https://doi.org/10.1021/jf030183j

CHA, Y. J., CADWALLADER, K. R., & BAEK, H. H. (1993). Volatile Flavor Components in Snow Crab Cooker Effluent and Effluent Concentrate. Journal of Food Science, 58(3), 525–530. https://doi.org/10.1111/j.1365-2621.1993.tb04316.x

Chiurugwi, T., Kemp, S., Powell, W., & Hickey, L. T. (2019). Speed breeding orphan crops. Theoretical and Applied Genetics, 132(3), 607–616. https://doi.org/10.1007/s00122-018-3202-7

Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G. W., … Huang, B. (2014). Erratum: Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system (Cell (2013)). Cell, 156(1–2), 373. https://doi.org/10.1016/j.cell.2013.12.033

Chen, X., Culbreath, A., Brenneman, T., Holbrook Jr, C., & Guo, B. (2008). Identification and cloning of TSWV resistance gene (s) in cultivated peanuts and development of markers for breeding selection. American Phytopathological Society Abstracts.

Chu, Y., Gill, R., Clevenger, J., Timper, P., Holbrook, C. C., & Ozias-Akins, P. (2016). Identification of Rare Recombinants Leads to Tightly Linked Markers for Nematode Resistance in Peanut. Peanut Science, 43(2), 88–93. https://doi.org/10.3146/ps16-12.1

Chu, Y., Holbrook, C. C., & Ozias-Akins, P. (2009). Two alleles of ahFAD2B control the high oleic acid trait in cultivated peanut. Crop Science, 49(6), 2029–2036. https://doi.org/10.2135/cropsci2009.01.0021

Chu, Y., Wu, C. L., Holbrook, C. C., Tillman, B. L., Person, G., & Ozias-Akins, P. (2011). Marker-Assisted Selection to Pyramid Nematode Resistance and the High Oleic Trait in Peanut. The Plant Genome Journal, 4(2), 110. https://doi.org/10.3835/plantgenome2011.01.0001

Chuan, T. W., Xin, D. Y., Dian, X. C., Shan, L. Y., Guang, Z. L., Yue, Y. T., & Jian, Z. X. (2007). Isolation of simple sequence repeats from groundnut. Electronic Journal of Biotechnology, 10(3), 473–480. https://doi.org/10.2225/vol10-issue3-fulltext-10

Collard, B. C. Y., & Mackill, D. J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences, 363(1491), 557–572. https://doi.org/10.1098/rstb.2007.2170

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica, 142(1–2), 169–196. https://doi.org/10.1007/s10681-005-1681-5

Cossetin, J. F., da Silva Brum, E., Casoti, R., Camponogara, C., Dornelles, R. C., Maziero, M., … Trevisan, G. (2019). Peanut leaf extract has antioxidant and anti-inflammatory activity but no acute toxic effects. Regulatory Toxicology and Pharmacology, 107(May). https://doi.org/10.1016/j.yrtph.2019.104407

Dash, S., Cannon, E. K. S., Kalberer, S. R., Farmer, A. D., & Cannon, S. B. (2016). PeanutBase and other bioinformatic resources for peanut. In Peanuts (pp. 241–252). Elsevier.

Deschamps, S., la Rota, M., Ratashak, J. P., Biddle, P., Thureen, D., Farmer, A., … Campbell, M. A. (2010). Rapid Genome-wide Single Nucleotide Polymorphism Discovery in Soybean and Rice via Deep Resequencing of Reduced Representation Libraries with the Illumina Genome Analyzer. The Plant Genome Journal, 3(1), 53. https://doi.org/10.3835/plantgenome2009.09.0026

De Paula, A. F., Dinato, N. B., Vigna, B. B. Z., & Fávero, A. P. (2017). Recombinants from the crosses between amphidiploid and cultivated peanut (Arachis hypogaea) for pest-resistance breeding programs. PLoS ONE, 12(4), 1–15. https://doi.org/10.1371/journal.pone.0175940

Drake, M. A., & Civille, G. V. (2003). Flavor lexicons. Comprehensive Reviews in Food Science and Food Safety, 2(1), 33–40. https://doi.org/10.1111/j.1541-4337.2003.tb00013.x

Drake, M. A., Gerard, P. D., Kleinhenz, J. P., & Harper, W. J. (2003). Application of an electronic nose to correlate with descriptive sensory analysis of aged Cheddar cheese. LWT - Food Science and Technology, 36(1), 13–20. https://doi.org/10.1016/S0023-6438(02)00216-5

Du, M., Li, N., Niu, B., Liu, Y., You, D., Jiang, D., … Wang, W. (2018). De novo transcriptome analysis of Bagarius yarrelli (Siluriformes: Sisoridae) and the search for potential SSR markers using RNA-Seq. PLoS ONE, 13(2), 1–11. https://doi.org/10.1371/journal.pone.0190343

Evans, J., Kim, J., Childs, K. L., Vaillancourt, B., Crisovan, E., Nandety, A., … Buell, C. R. (2014). Nucleotide polymorphism and copy number variant detection using exome capture and next-generation sequencing in the polyploid grass Panicum virgatum. Plant Journal, 79(6), 993–1008. https://doi.org/10.1111/tpj.12601

FAO. Food and Agricultural Organization of the United Nation, FAO Statistical Database http://faostat.fao.org/faostat/collections? subset=agriculture); 2009.

FAOSTAT. FAO statistics division 2014.http://faostat.fao.org/ 2014.

Flint-Garcia, S. A., Thornsberry, J. M., & Buckler, E. S. (2003). Structure of Linkage Disequilibrium in Plants. Annual Review of Plant Biology, 54(1), 357–374. https://doi.org/10.1146/annurev.arplant.54.031902.134907

Fontana Capalbo, D. M., Valicente, F. H., Moraes, I. D. O., & Pelizer, L. H. (2001). Solid-state fermentation of Bacillus thuringiensis tolworthi to control fall armyworm in maize. Electronic Journal of Biotechnology, 4(2), 112–115. https://doi.org/10.2225/vol4-issue2-fulltext-5

Froenicke, L., Nwosu, V., Michelmore, R., Ozias-Akins, P., Scheffler, B., & Wilson, R. F. (n.d.). Issue: Discussion of Component Strategies for Peanut Genome Mapping.

Galli, A., Svegliati-Baroni, G., Ceni, E., Milani, S., Ridolfi, F., Salzano, R., … Casini, A. (2005). Oxidative stress stimulates proliferation and invasiveness of hepatic stellate cells via a MMP2-mediated mechanism. Hepatology, 41(5), 1074–1084. https://doi.org/10.1002/hep.20683

Gautami, B., Pandey, M. K., Vadez, V., Nigam, S. N., Ratnakumar, P., Krishnamurthy, L., … Hoisington, D. A. (2012). QTL analysis and consensus genetic map for drought tolerance traits based on three RIL populations of cultivated groundnut (*Arachis hypogaea* L.). Mol Breed, 32, 757–772.

Gautami, B., Ravi, K., Narasu, M. L., Hoisington, D. A., & Varshney, R. K. (2009). Novel set of groundnut SSR markers for germplasm analysis and interspecific transferability. International Journal of Integrative Biology, 7(2), 100–106.

Grimm, D. T., Sanders, T. H., Pattee, H. E., Williams, D. E., & Sanchez-Dominguez, S. (1996). Chemical Composition of *Arachis hypogaea* L. Subsp. *hypogaea* Var. hirsuta Peanuts 1 . Peanut Science, 23(2), 111–116. https://doi.org/10.3146/i0095-3679-23-2-9

Guimaraes, P. M., Guimaraes, L. A., Morgante, C. V., Silva, O. B., Araujo, A. C. G., Martins, A. C. Q., … Brasileiro, A. C. M. (2015). Root transcriptome analysis of wild peanut reveals candidate genes for nematode resistance. PLoS ONE, 10(10), 1–22. https://doi.org/10.1371/journal.pone.0140937

Guo, B., Fedorova, N. D., Chen, X., Wan, C. H., Wang, W., Nierman, W. C., … Yu, J. (2011). Gene expression profiling and identification of resistance genes to Aspergillus flavus infection in peanut through EST and microarray strategies. Toxins, 3(7), 737–753. https://doi.org/10.3390/toxins3070737

Gupta, P. K., Kulwal, P. L., & Jaiswal, V. (2019). Association mapping in plants in the post-GWAS genomics era. Advances in Genetics, *104*, 75–154.

Guo, Y., Abernathy, B., Zeng, Y., & Ozias-Akins, P. (2015). TILLING by sequencing to identify induced mutations in stress resistance genes of peanut (*Arachis hypogaea*). BMC Genomics, 16(1), 1–13. https://doi.org/10.1186/s12864-015-1348-0

Gupta, P. K., Kumar, J., Mir, R. R., & Kumar, A. (2010). Edited by Jules Janick Cleaved amplified polymorphism. 145–218.

Hashim, I. B., Koehler, P. E., & Eitenmiller, R. R. (1993). Tocopherols in runner and virginia peanut cultivars at various maturity stages. Journal of the American Oil Chemists' Society, 70(6), 633–635. https://doi.org/10.1007/BF02545333

He, G., & Prakash, C. (2001). Evaluation of genetic relationships among botanical varieties of cultivated peanut (*Arachis hypogaea* L.) using AFLP markers. Genetic Resources and Crop Evolution, 48(4), 347–352. https://doi.org/10.1023/A:1012019600318

He, G., Meng, R., Newman, M., Gao, G., Pittman, R. N., & Prakash, C. S. (2003). Microsatellites as DNA markers in cultivated peanut (*Arachis hypogaea* L.). BMC Plant Biology, 3, 1–6. https://doi.org/10.1186/1471-2229-3-3

Herselman, L., Thwaites, R., Kimmins, F. M., Courtois, B., Van Der Merwe, P. J. A., & Seal, S. E. (2004). Identification and mapping of AFLP markers linked to peanut (*Arachis hypogaea* L.) resistance to the aphid vector of groundnut rosette disease. Theoretical and Applied Genetics, 109(7), 1426–1433. https://doi.org/10.1007/s00122-004-1756-z

Hinds, M. J. (1995). Fatty acid composition of Caribbean-grown peanuts (*Arachis hypogaea* L.) at three maturity stages. Food Chemistry, 53(1), 7–14. https://doi.org/10.1016/0308-8146(95)95779-6

Holbrook, C. C., & Dong, W. (2005). Development and evaluation of a Mini-core collection for the U.S. peanut germplasm collection. Crop Science, 45(4), 1540–1544. https://doi.org/10.2135/cropsci2004.0368

Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell, 157(6), 1262–1278. https://doi.org/10.1016/j.cell.2014.05.010

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., … Han, B. (2010). Genome-wide asociation studies of 14 agronomic traits in rice landraces. Nature Genetics, 42(11), 961–967. https://doi.org/10.1038/ng.695

Huang, X., Zhao, Y., Wei, X., Li, C., Wang, A., Zhao, Q., … Han, B. (2012). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nature Genetics, 44(1), 32–39. https://doi.org/10.1038/ng.1018

Ibitoye, D. O., & Akin-Idowu, P. E. (2010). Marker-assisted-selection (MAS): A fast track to increase genetic gain in horticultural crop breeding. African Journal of Biotechnology, 9(52), 8889–8895.

Isleib, T. G., Pattee, H. E., & Giesbrecht, F. G. (1995). Ancestral Contributions to Roasted Peanut Attribute 1 . Peanut Science, 22(1), 42–48. https://doi.org/10.3146/pnut.22.1.0009

Isleib, T. G., Wilson, R. F., & Novitzky, W. P. (2006). Partial dominance, pleiotropism, and epistasis in the inheritance of the high-oleate trait in peanut. Crop Science, 46(3), 1331–1335. https://doi.org/10.2135/cropsci2005.09-0313

Ivandic, V., Thomas, W. T. B., Nevo, E., Zhang, Z., & Forster, B. P. (2003). Associations of simple sequence repeats with quantitative trait variation including biotic and abiotic stress tolerance in Hordeum spontaneum. Plant Breeding, 122(4), 300–304. https://doi.org/10.1046/j.1439-0523.2003.00841.x

Janila, P., Pandey, M. K., Shasidhar, Y., Variath, M. T., Sriswathi, M., Khera, P., … Varshney, R. K. (2016). Molecular breeding for introgression of fatty acid desaturase mutant alleles (ahFAD2A and ahFAD2B) enhances oil quality in high and low oil containing peanut genotypes. Plant Science, 242, 203–213. https://doi.org/10.1016/j.plantsci.2015.08.013

Jannink, J. L., & Walsh, B. (2002). Quantitative Genetics, Genomics and Plant Breeding.

Jia, G., Huang, X., Zhi, H., Zhao, Y., Zhao, Q., Li, W., … Han, B. (2013). A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (Setaria italica). Nature Genetics, 45(8), 957–961. https://doi.org/10.1038/ng.2673

Jordan, B. S., Culbreath, A. K., Brenneman, T. B., Kemerait, R. C., & Branch, W. D. (2017). Late leaf spot severity and yield of new peanut breeding lines and cultivars grown without fungicides. Plant Disease, 101(11), 1843–1850. https://doi.org/10.1094/PDIS-02-17-0165-RE

Jonnala, R. S., Dunford, N. T., & Dashiell, K. E. (2006). Tocopherol, phytosterol and phospholipid compositions of new high oleic peanut cultivars. Journal of Food Composition and Analysis, 19(6–7), 601–605. https://doi.org/10.1016/j.jfca.2006.01.005

Kantartzi, S. K., & Stewart, J. M. D. (2008). Association analysis of fibre traits in Gossypium arboreum accessions. Plant Breeding, 127(2), 173–179. https://doi.org/10.1111/j.1439-0523.2008.01490.x

Kaya, C., Hamamci, C., Baysal, A., Akba, O., Erdogan, S., & Saydut, A. (2009). Methyl ester of peanut (*Arachis* hypogea L.) seed oil as a potential feedstock for biodiesel production. Renewable Energy, 34(5), 1257–1260. https://doi.org/10.1016/j.renene.2008.10.002

Knauft, D. A., Moore, K. M., & Gorbet, D. W. (1993). Further Studies On The Inheritance Of Fatty Acid Composition In Peanut 1 . Peanut Science, 20(2), 74–76. https://doi.org/10.3146/i0095-3679-20-2-2

Kumar, L. S. (1999). DNA markers in plant improvement: An overview. Biotechnology Advances, 17(2–3), 143–182. https://doi.org/10.1016/S0734-9750(98)00018-4

Kump, K. L., Bradbury, P. J., Wisser, R. J., Buckler, E. S., Belcher, A. R., Oropeza-Rosas, M. A., … Holland, J. B. (2011). Genome-wide association study of quantitative resistance to

southern leaf blight in the maize nested association mapping population. Nature Genetics, 43(2), 163–168. https://doi.org/10.1038/ng.747

Lang, W., Cai, J., Chen, F., Zhu, J., Hu, X., & Zhong, J. (2019). RNA-Seq analyses demonstrate EVI-1–induced morbid hematopoiesis and developmental abnormality in zebrafish were related with MAPK pathway. Hematological Oncology, 37(3), 326–329. https://doi.org/10.1002/hon.2609

Lee, J. M., Pae, S.-B., Choung, M.-G., Lee, M.-H., Kim, S.-U., Oh, E., … Oh, I. S. (2016). Determination of fatty acid composition in peanut seed by near infrared reflectance spectroscopy. 61(1), 64–69.

Li, Y., Chen, C. Y., Knapp, S. J., Culbreath, A. K., Holbrook, C. C., & Guo, B. (2011). Characterization of Simple Sequence Repeat (SSR) Markers and Genetic Relationships within Cultivated Peanut (*Arachis hypogaea* L.). Peanut Science, 38(1), 1–10. https://doi.org/10.3146/ps10-10.1

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., … Zhang, Z. (2012). GAPIT: Genome association and prediction integrated tool. Bioinformatics, 28(18), 2397–2399. https://doi.org/10.1093/bioinformatics/bts444

Lippert, C., Listgarten, J., Liu, Y., Kadie, C. M., Davidson, R. I., & Heckerman, D. (2011). FaST linear mixed models for genome-wide association studies. Nature Methods, 8(10), 833.

Liu, L., Dang, P. M., & Chen, C. Y. (2015). Development and Utilization of InDel Markers to Identify Peanut (Arachis hypogaea) Disease Resistance . Frontiers in Plant Science , Vol. 6, p. 988. Retrieved from https://www.frontiersin.org/article/10.3389/fpls.2015.00988

Liu, R., Jin, Q., Huang, J., Liu, Y., Wang, X., Mao, W., & Wang, S. (2011). Photodegradation of Aflatoxin B1 in peanut oil. European Food Research and Technology, 232(5), 843–849. https://doi.org/10.1007/s00217-011-1452-6

Long, Q., Rabanal, F. A., Meng, D., Huber, C. D., Farlow, A., Platzer, A., … Nordborg, M. (2013). Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden. Nature Genetics, 45(8), 884–890. https://doi.org/10.1038/ng.2678

Lykomitros, D., Fogliano, V., & Capuano, E. (2016). Flavor of roasted peanuts (*Arachis hypogaea*) — Part II: Correlation of volatile compounds to sensory characteristics. Food Research International, 89, 870–881. https://doi.org/10.1016/j.foodres.2016.08.01

Luo, L., Xia, H., & Lu, B. (2019). Crop breeding for drought resistance. *Frontiers in Plant Science*, *10*, 314.7

Ma, Y., Zhang, L., & Huang, X. (2014). Genome modification by CRISPR/Cas9. The FEBS Journal, 281(23), 5186–5193.

Mackay, I., & Powell, W. (2007). Methods for linkage disequilibrium mapping in crops. Trends in Plant Science, 12(2), 57–63. https://doi.org/10.1016/j.tplants.2006.12.001

Majewski, J., & Pastinen, T. (2011). The study of eQTL variations by RNA-seq: From SNPs to phenotypes. Trends in Genetics, 27(2), 72–79. https://doi.org/10.1016/j.tig.2010.10.006

Malosetti, M., Van Der Linden, C. G., Vosman, B., & Van Eeuwijk, F. A. (2007). A mixed-model approach to association mapping using pedigree information with an illustration of resistance to Phytophthora infestans in potato. Genetics, 175(2), 879–889. https://doi.org/10.1534/genetics.105.054932

Mardis, E. R. (2013). Next-Generation Sequencing Platforms. Annual Review of Analytical Chemistry, 6(1), 287–303. https://doi.org/10.1146/annurev-anchem-062012-092628

Mathew, B., Sillanpää, M. J., & Léon, J. (2019). Advances in statistical methods to handle large data sets for genome-wide association mapping in crop breeding. 437–450. https://doi.org/10.19103/as.2019.0051.20

McGorrin, R. J. (2002). Character impact compounds: Flavors and off-flavors in foods. FOOD SCIENCE AND TECHNOLOGY-NEW YORK-MARCEL DEKKER-, 375–413.

McMeans, J. L., Sanders, T. H., Wood, B. W., & Blankenship, P. D. (1990). Soil Temperature Effects on Free Carbohydrate Concentrations in Peanut (*Arachis hypogaea* L.) Seed 1 . Peanut Science, 17(1), 31–35. https://doi.org/10.3146/i0095-3679-17-1-11

Meilgaard, M. C. (1991). Current Progress in Sensory Analysis. A Review. Journal of the American Society of Brewing Chemists, 49(3), 101–109. https://doi.org/10.1094/asbcj-49-0101

Mizuno, K., Tanaka, M., Nozaki, S., Mizuma, H., Ataka, S., Tahara, T., … Watanabe, Y. (2008). Antifatigue effects of coenzyme Q10 during physical fatigue. Nutrition, 24(4), 293–299. https://doi.org/10.1016/j.nut.2007.12.007

Moose, S. P., & Mumm, R. H. (2008). Molecular plant breeding as the foundation for 21st century crop improvement. Plant Physiology, 147(3), 969–977. https://doi.org/10.1104/pp.108.118232

Mondal, S., & Badigannavar, A. M. (2015). Peanut rust (Puccinia arachidis Speg.) disease: its background and recent accomplishments towards disease resistance breeding. Protoplasma, 252(6), 1409–1420. https://doi.org/10.1007/s00709-015-0783-8

Morris, G. P., Ramu, P., Deshpande, S. P., Hash, C. T., Shah, T., Upadhyaya, H. D., … Kresovich, S. (2013). Population genomic and genome-wide association studies of agroclimatic traits in

sorghum. Proceedings of the National Academy of Sciences of the United States of America, 110(2), 453–458. https://doi.org/10.1073/pnas.1215985110

Nagy, E., Guo, S., Khanal, S., Taylor, C., Ozias-Akins, P., Stalker, H. T., & Nielsen, N. (2010). Developing a high density molecular map of the A-genome species A. *duranensis*. Proc. American Peanut Research and Education Society (APRES), 12–15.

Nawade, B., Mishra, G. P., Radhakrishnan, T., Dodia, S. M., Ahmad, S., Kumar, A., … Kundu, R. (2018). High oleic peanut breeding: Achievements, perspectives, and prospects. Trends in Food Science and Technology, 78(June 2017), 107–119. https://doi.org/10.1016/j.tifs.2018.05.022

Neta, E. R., Sanders, T., & Drake, M. A. (2010). Understanding peanut flavor: a current review. Handbook of Fruit and Vegetable Flavors, 985.

Neya, F., Koita, K., Elise, S., Zagre, B., Nana, A., Burow, M., & Sankara, P. (2018). Agro-morphological Diversity of Six Peanut (Arachis hypogaea L.) Breeding Lines from Three Geographical Areas. Journal of Experimental Agriculture International, 19(5), 1–12. https://doi.org/10.9734/jeai/2017/37686

Nichols, J. A., & Moller, T. H. (2005). International cooperation in oil spill response. 2005 International Oil Spill Conference, IOSC 2005, (June), 4103. https://doi.org/10.7901/2169-3358-1991-1-61

Nursten H. E. (2005) The Maillard Reaction: Chemistry, Biochemistry, and Implications. books.google.com.

Oliver, J. (2013). 済無No Title No Title. In Journal of Chemical Information and Modeling (Vol. 53). https://doi.org/10.1017/CBO9781107415324.004

Olsen, K. M., & Purugganan, M. D. (2002). Molecular evidence on the origin and evolution of glutinous rice. Genetics, 162(2), 941–950.

Pandey, M. K., Gautami, B., Jayakumar, T., Sriswathi, M., Upadhyaya, H. D., Gowda, M. V. C., … Varshney, R. K. (2012). Highly informative genic and genomic SSR markers to facilitate molecular breeding in cultivated groundnut (*Arachis hypogaea*). Plant Breeding, 131(1), 139–147. https://doi.org/10.1111/j.1439-0523.2011.01911.x

Pandey, M. K., Wang, M. L., Qiao, L., Feng, S., Khera, P., Wang, H., … Guo, B. (2014). Identification of QTLs associated with oil content and mapping FAD2 genes and their relative contribution to oil quality in peanut (*Arachis hypogaea* L.). BMC Genetics, 15(1), 1–14. https://doi.org/10.1186/s12863-014-0133-4

Pattee, H. E., & Giesbrecht, F. G. (1990). Roasted Peanut Flavor Variation Across Germplasm Sources 1 . Peanut Science, 17(2), 109–112. https://doi.org/10.3146/i0095-3679-17-2-15

Pattee, H. E., Giesbrecht, F. G., & Isleib, T. G. (1995). Roasted Peanut Flavor Intensity Variations Among U.S. Genotypes 1 . Peanut Science, 22(2), 158–162. https://doi.org/10.3146/i0095-3679-22-2-16

Pattee, H. E., Isleib, T. G., & Giesbrecht, F. G. (1998). Variation in Intensity of Sweet and Bitter Sensory Attributes Across Peanut Genotypes 1 . Peanut Science, 25(2), 63–69. https://doi.org/10.3146/i0095-3679-25-2-2

Pattee, H. E., Isleib, T. G., Giesbrecht, F. G., & McFeeters, R. F. (2000). Investigations into genotypic variations of peanut carbohydrates. Journal of Agricultural and Food Chemistry, 48(3), 750–756. https://doi.org/10.1021/jf9910739

Pattee, H. E., Isleib, T. G., Giesbrecht, F. G., & McFeeters, R. F. (2000). Relationships of sweet, bitter, and roasted peanut sensory attributes with carbohydrate components in peanuts.

Journal of Agricultural and Food Chemistry, 48(3), 757–763. https://doi.org/10.1021/jf9910741

Plundrich, N. J., White, B. L., Dean, L. L., Davis, J. P., Foegeding, E. A., & Lila, M. A. (2015). Stability and immunogenicity of hypoallergenic peanut protein-polyphenol complexes during in vitro pepsin digestion. Food and Function, 6(7), 2145–2154. https://doi.org/10.1039/c5fo00162e

Qu, S., Yang, X., Li, X., Wang, J., Gao, Y., Shang, R., … Li, H. (2015). Circular RNA: A new star of noncoding RNAs. Cancer Letters, 365(2), 141–148. https://doi.org/10.1016/j.canlet.2015.06.003

Ravi, K., Vadez, V., Isobe, S., Mir, R. R., Guo, Y., Nigam, S. N., … Knapp, S. J. (2011). Identification of several small-effect main QTLs and large number of epistatic QTLs for drought tolerance in groundnut (*Arachis hypogaea* L.). Theor. Appl. Genet, 122, 1119–1132.

Ribaut, J. M., & Ragot, M. (2007). Marker-assisted selection to improve drought adaptation in maize: The backcross approach, perspectives, limitations, and alternatives. Journal of Experimental Botany, 58(2), 351–360. https://doi.org/10.1093/jxb/erl214

Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases. Science, 273(5281), 1516–1517.

Rothman, N., Garcia-Closas, M., Chatterjee, N., Malats, N., Wu, X., Figueroa, J. D., … Chanock, S. J. (2010). A multi-stage genome-wide association study of bladder cancer identifies multiple susceptibility loci. Nature Genetics, 42(11), 978–984. https://doi.org/10.1038/ng.687

Sander, J. D., & Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nature Biotechnology, 32(4), 347–350. https://doi.org/10.1038/nbt.2842

Sanders, T. H., Pattee, H. E., Vercellotti, J. R., & Bett, K. L. (1995). Advances in peanut flavor quality. Advances in Peanut Science, 528–553.

Sarvamangala, C., Gowda, M. V. C., & Varshney, R. K. (2011). Identification of quantitative trait loci for protein content, oil content and oil quality for groundnut (*Arachis hypogaea* L.). Field Crops Research, 122(1), 49–59. https://doi.org/10.1016/j.fcr.2011.02.010

Schafer, A. J., & Hawkins, J. R. (1998). hun1an genetics.

Schirack, A. V., Drake, M., Sanders, T. H., & Sandeep, K. P. (2006). Impact of microwave blanching on the flavor of roasted peanuts. Journal of Sensory Studies, 21(4), 428–440. https://doi.org/10.1111/j.1745-459X.2006.00075.x

Scholten, H. (1994). Plant breeding reviews, volume 10. In Scientia Horticulturae (Vol. 57). https://doi.org/10.1016/0304-4238(94)90045-0

Seijo, J. G., Lavia, G. I., Fernández, A., Krapovickas, A., Ducasse, D., & Moscone, E. A. (2004). Physical mapping of the 5S and 18S-25S rRNA genes by FISH as evidence that *Arachis duranensis* and A. *ipaensis* are the wild diploid progenitors of A. *hypogaea* (Leguminosae). American Journal of Botany, 91(9), 1294–1303. https://doi.org/10.3732/ajb.91.9.1294

Seren, Ü., Vilhjálmsson, B. J., Horton, M. W., Meng, D., Forai, P., Huang, Y. S., … Nordborg, M. (2013). GWAPP: A web application for genome-Wide association mapping in arabidopsis. Plant Cell, 24(12), 4793–4805. https://doi.org/10.1105/tpc.112.108068

Shalek, A. K., Satija, R., Shuga, J., Trombetta, J. J., Gennert, D., Lu, D., … Regev, A. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature, 510(7505), 363–369. https://doi.org/10.1038/nature13437

Shasidhar, Y., Vishwakarma, M. K., Pandey, M. K., Janila, P., Variath, M. T., Manohar, S. S., … Varshney, R. K. (2017). Molecular mapping of oil content and fatty acids using dense genetic

maps in groundnut (*Arachis hypogaea* L.). Frontiers in Plant Science, 8. https://doi.org/10.3389/fpls.2017.00794

Shiriki, D., Igyor, M. A., & Gernah, D. I. (2015). Nutritional Evaluation of Complementary Food Formulations from Maize, Soybean and Peanut Fortified with <i>Moringa oleifera</i> Leaf Powder. Food and Nutrition Sciences, 06(05), 494–500. https://doi.org/10.4236/fns.2015.65051

Singh Kesawat, M., & Das, B. K. (2009). Molecular Markers: It's Application in Crop Improvement Types of Molecular Markers. J. Crop Sci. Biotech, 12(4), 169–181. https://doi.org/10.1007/s12892-009-0124-6

Singh, K. P., Raina, S. N., & Singh, A. K. (1996). Variation in chromosomal DNA associated with the evolution of *Arachis* species. Genome, 39(5), 890–897. https://doi.org/10.1139/g96-112

Skøt, L., Humphreys, J., Humphreys, M. O., Thorogood, D., Gallagher, J., Sanderson, R., … Thomas, I. D. (2007). Association of candidate genes with flowering time and water-soluble carbohydrate content in Lolium perenne (L.). Genetics, 177(1), 535–547. https://doi.org/10.1534/genetics.107.071522

Skøt, L., Humphreys, M. O., Armstead, I., Heywood, S., Skøt, K. P., Sanderson, R., … Hamilton, N. R. S. (2005). An association mapping approach to identify flowering time genes in natural populations of Lolium perenne (L.). Molecular Breeding, 15(3), 233–245. https://doi.org/10.1007/s11032-004-4824-9

Somers, D. J., Langridge, P., & Gustafson, J. P. (n.d.). Springer Protocols Methods in Molecular Biology 513 Plant Genomics Methods and Protocols.

Song, G. Q., Li, M. J., Xiao, H., Wang, X. J., Tang, R. H., Xia, H., … Bi, Y. P. (2010). EST sequencing and SSR marker development from cultivated peanut (*Arachis hypogaea* L.). Electronic Journal of Biotechnology, 13(3), 1–9. https://doi.org/10.2225/vol13-issue3-fulltext-10

Song, H., Wang, P., Lin, J. Y., Zhao, C., Bi, Y., & Wang, X. (2016). Genome-wide identification and characterization of WRKY gene family in peanut. Frontiers in Plant Science, 7(APR2016), 1–15. https://doi.org/10.3389/fpls.2016.00534

Song, H., Wang, P., Lin, J.-Y., Zhao, C., Bi, Y., & Wang, X. (2016). Genome-Wide Identification and Characterization of WRKY Gene Family in Peanut. Frontiers in Plant Science, 7. https://doi.org/10.3389/fpls.2016.00534

Stalker, H. T., & Simpson, C. E. (1995). Genetic resources in *Arachis*. Advances in Peanut Science, 14–53.

Stalker, H. T., Dhesi, J. S., Parry, D. C., & Hahn, J. H. (1991). Cytological and Interfertility Relationships of *Arachis* Section *Arachis*. American Journal of Botany, 78(2), 238. https://doi.org/10.2307/2445247

Stich, B., Melchinger, A. E., Heckenberger, M., Möhring, J., Schechert, A., & Piepho, H. P. (2008). Association mapping in multiple segregating populations of sugar beet (Beta *vulgaris* L.). Theoretical and Applied Genetics, 117(7), 1167–1179. https://doi.org/10.1007/s00122-008-0854-8

Subramanian, V., Gurtu, S., Nageswara Rao, R. C., & Nigam, S. N. (2000). Identification of DNA polymorphism in cultivated groundnut using random amplified polymorphic DNA (RAPD) assay. Genome, 43(4), 656–660. https://doi.org/10.1139/g00-034

Talcott, S. T., Passeretti, S., Duncan, C. E., & Gorbet, D. W. (2005). Polyphenolic content and sensory properties of normal and high oleic acid peanuts. Food Chemistry, 90(3), 379–388. https://doi.org/10.1016/j.foodchem.2004.04.011

Tester, Mark and Langridge, P. (2010). Breeding Technologies to Increase. Science, 818(5967), 818–822. https://doi.org/10.1126/science.1183700

Thornsberry, J. M., Goodman, M. M., Doebley, J., Kresovich, S., Nielsen, D., & Buckler, E. S. (2001). Dwarf8 polymorphisms associate with variation in flowering time. Nature Genetics, 28(3), 286–289. https://doi.org/10.1038/90135

Tillman, B. L., & McKinney, J. L. (2018). Relationships among symptoms of spotted wilt disease of peanut and their potential impact on crop productivity and resistance breeding. Plant Breeding, 137(5), 757–762. https://doi.org/10.1111/pbr.12638

Upadhyaya, H. D., Bramel, P. J., Ortiz, R., & Singh, S. (2002). Developing a Mini-core of peanut for utilization of genetic resources. Crop Science, 42(6), 2150–2156. https://doi.org/10.2135/cropsci2002.2150

Upadhyaya, H. D., Dwivedi, S. L., Nadaf, H. L., & Singh, S. (2011). Phenotypic diversity and identification of wild *Arachis* accessions with useful agronomic and nutritional traits. Euphytica, 182(1), 103–115. https://doi.org/10.1007/s10681-011-0518-7

Van Verk, M. C., Hickman, R., Pieterse, C. M. J., & Van Wees, S. C. M. (2013). RNA-Seq: Revelation of the messengers. Trends in Plant Science, 18(4), 175–179. https://doi.org/10.1016/j.tplants.2013.02.001

Varshney, R. K., & Pandey, M. K. (n.d.). The Peanut Genome Compendium of Plant Genomes Series Editor: Chittaranjan Kole. Retrieved from http://www.springer.com/series/11805

Varshney, R. K., Hoisington, D. A., & Tyagi, A. K. (2006). Advances in cereal genomics and applications in crop breeding. Trends in Biotechnology, 24(11), 490–499. https://doi.org/10.1016/j.tibtech.2006.08.006

Varshney, R. K., Nayak, S. N., May, G. D., & Jackson, S. A. (2009). Next-generation sequencing technologies and their implications for crop genetics and breeding. Trends in Biotechnology, 27(9), 522–530. https://doi.org/10.1016/j.tibtech.2009.05.006

Wang, H., Khera, P., Huang, B., Yuan, M., Katam, R., Zhuang, W., … Guo, B. (2016). Analysis of genetic diversity and population structure of peanut cultivars and breeding lines from China, India and the US using simple sequence repeat markers. Journal of Integrative Plant Biology, 58(5), 452–465. https://doi.org/10.1111/jipb.12380

Wang, M. L., Barkley, N. A., Chen, Z., & Pittman, R. N. (2011). FAD2 gene mutations significantly alter fatty acid profiles in cultivated peanuts (*Arachis hypogaea*). Biochemical Genetics, 49(11–12), 748–759. https://doi.org/10.1007/s10528-011-9447-3

Wang, M. L., Sukumaran, S., Barkley, N. A., Chen, Z., Chen, C. Y., Guo, B., … Yu, J. (2011). Population structure and marker-trait association analysis of the US peanut (*Arachis hypogaea* L.) Mini-core collection. Theoretical and Applied Genetics, 123(8), 1307–1317. https://doi.org/10.1007/s00122-011-1668-7

Wright, G. C., Borgognone, M. G., Connor, D. J. O., Rachaputi, R. C. N., Henry, R. J., Furtado, A., … Freischfresser, D. B. (2018). Breeding for improved blanchability in peanut: Phenotyping, genotype × environment interaction and selection. Crop and Pasture Science, 69(12), 1237–1250. https://doi.org/10.1071/CP18156

Xu, Y., & Crouch, J. H. (2008). Marker-assisted selection in plant breeding: From publications to practice. Crop Science, 48(2), 391–407. https://doi.org/10.2135/cropsci2007.04.0191

Young, C. T. (1996). Peanut oil. Bailey's Industrial Oil and Fat Products, 3, 377–392.

Yuan, M., Gong, L., Meng, R., Li, S., Dang, P., Guo, B., & He, G. (2010). Development of trinucleotide (GGC)n SSR markers in peanut (*Arachis hypogaea* L.). Electronic Journal of Biotechnology, 13(6). https://doi.org/10.2225/vol13-issue6-fulltext-6

Yuan, M., Zhu, J., Gong, L., He, L., Lee, C., Han, S., … He, G. (2019). Mutagenesis of FAD2 genes in peanut with CRISPR/Cas9 based gene editing. BMC Biotechnology, 19(1), 1–7. https://doi.org/10.1186/s12896-019-0516-8

Zeng, L., Meredith, W. R., Gutiérrez, O. A., & Boykin, D. L. (2009). Identification of associations between SSR markers and fiber traits in an exotic germplasm derived from multiple crosses among Gossypium tetraploid species. Theoretical and Applied Genetics, 119(1), 93–103. https://doi.org/10.1007/s00122-009-1020-7

Zhang, H., Shi, C., Wu, J., Ren, Y., Li, C., Zhang, D., & Zhang, Y. (2004). Analysis of genetic and genotype x environment interaction effects from embryo, cytoplasm and maternal plant for oleic acid content of Brassica napus L. Plant Science, 167(1), 43–48. https://doi.org/10.1016/j.plantsci.2004.02.023

Zhao, K., Tung, C. W., Eizenga, G. C., Wright, M. H., Ali, M. L., Price, A. H., … McCouch, S. R. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in Oryza sativa. Nature Communications, 2(1), 1–10. https://doi.org/10.1038/ncomms1467

Zhao, T., Fu, Y., Zhu, J., Liu, Y., Zhang, Q., Yi, Z., … Deng, H. (2018). Single-Cell RNA-Seq Reveals Dynamic Early Embryonic-like Programs during Chemical Reprogramming. Cell Stem Cell, 23(1), 31-45.e7. https://doi.org/10.1016/j.stem.2018.05.025

Zhao, Y., Zhang, C., Chen, H., Yuan, M., Nipper, R., Prakash, C. S., … He, G. (2016). QTL mapping for bacterial wilt resistance in peanut (Arachis hypogaea L.). Molecular Breeding, 36(2), 1–11. https://doi.org/10.1007/s11032-015-0432-0

Zhong, R., Zhou, M., Zhao, C., Hou, L., Li, C., Wang, X., … Xia, H. (2016). SSR marker development from peanut gynophore transcriptome sequencing. Plant Breeding, 135(1), 111–117. https://doi.org/10.1111/pbr.12336

Chapter Two

Genome-Wide Association Study of Sensory and Quality Attributes in Cultivated Peanut

2.1 Abstract

As more consumers are paying attention to food quality and nutrients, obtaining desirable flavor and quality attributes become an vital breeding objective for peanut. Currently, 120 accessions of the U.S. peanut Mini-core collection are used to investigate the genetic population structure and its potential usefulness on genetic marker- agronomic trait relationship through genome-wide association studies (GWAS). In order to analyze the flavor and quality attributes (including sensory attributes, sugars, fatty acids, and tocopherols), this study first used the Affymetrix version 2.0 SNP array to identify a total of 17,232 high-quality single nucleotide polymorphisms (SNPs) from the Mini-core Collection. 310, 474, 569, and 230 QTLs were identified related to sensory attributes, sugars, fatty acids, and tocopherols, respectively, including 14, 87, 105, and 73 significant QTLs for each profile. Among the significant QTLs, there were 4, 55, 86, and 49 QTLs with PVE higher than 15% for each profile, respectively (Table 3), indicating the high heritability of these traits. Consequently, genes were screened within 1 Mb of the significant QTLs, and a total of 136 candidate genes were identified functionally associated with corresponding traits, including 13 genes associated with sensory attributes, 65 genes with sugars, 35 genes with fatty acids and 23 genes with tocopherols, respectively. Among these candidate genes, most of them are known to be involved in the biosynthesis and metabolism process of carbohydrates, fatty acids, and polyphenols. These studies provided a prospective insight into the intricate genetic structure on sensory and quality attributes in peanut and would promote the marker-assisted selection (MAS) in peanut breeding with seed flavor and nutrition quality.

Keywords: peanut, marker-assisted selection (MAS), single-nucleotide polymorphisms (SNPs), genome-wide association studies (GWAS), sensory attributes, sugars, fatty acids, tocopherols

2.2 Introduction

Peanut (*Arachis hypogaea* L.), of the legume family, is one of the most important economic crops grown worldwide. Unlike other crops, a large part of the peanut production is used for human consumption, such as toasted peanut, peanut butter, savory snacks, oil production, and peanut flour (Kaya et al., 2009; Chang et al., 2013). Hence, how to obtain desirable flavor and quality attributes becomes a vital breeding objective for peanut.

Although there are extensive researches about chemical components of cultivated peanut, how the chemicals or processing steps affect the ultimate peanut flavor and nutrition has not been completely elucidated. Nursten in 2005 gave a well-accepted explanation that when under heat, foods with roast flavor will have the Maillard reaction, in which the polymerization of free amino acids with sugar can change the flavor. Also, data showed that the processing is crucial in producing flavor, besides the outside environment and chemicals inside the peanut. (Lykomitros et al., 2016). Hence, both raw materials and processing methods can change the flavor and nutrient of peanuts. Based on the previous research, the peanut sensory attributes have high inheritability in peanut, such as attribute sweet, bitter, and sweet aromatic, while quality attributes have even higher (Pattee et al., 1995, 2000).

Besides sensory attributes, the seed nutrient quality attributes, such as sugars, fatty acids, and tocopherols, are also as common study subjects due to their closely relation with human health, also due to their high-impact on peanut flavor (Schirack et al, 2006; Leclercq et al., 2007; Liu et al., 2011). Previous studies show that genotypic variation accounts for 38%-78% of the overall

variation of the known Carbohydrates, implying that high broad-sense heritability for such compositions. (Pattee, 2000). While the levels of tocopherols have been reported to be environmentally influenced, there may be some genetic forces in effect. (Baydar et al., 2005; Jonnala et al., 2006). Therefore, with these high hereditary characters, it could be more efficiently and effectively to conduct the peanut breeding for sensory and quality attributes by obtaining genetic molecular markers associated to corresponding traits and utilizing them in Marker-assisted selection, comparing to the long-term and high-cost traditional breeding strategy (Collard et al., 2007; Xu et al., 2008).

Even though some agronomic traits and diseases related QTLs and genes have already been mapped, the genetics study for sensory and quality attributes is primarily limited, due to the lack of applicable genetic markers and the comparatively large genome size (~2.7 Gb) and complex of the allotetraploid genome structure in cultivated peanuts (Bertioli et al, 2019). Nevertheless, the developed sequencing technologies made the whole-genome sequencing more straightforward and allowed the analysis of the impact of sensory and quality attributes on genomic variation peanut (Tseng et al., 2016; Wilson et al., 2016; Bertioli et al., 2016), such as specific-locus amplified fragment sequencing technology, which is an effective method for large-scale single nucleotide polymorphism (SNP) identification and genotyping (Sun et al., 2013).

Our objectives are to locate the genomic loci associated with the four major attributes, and screen candidate genes residing within 1 Mb of the identified QTLs. This study would provide the clue to understand the genetic mechanisms related to sensory and quality attributes and promote the marker-assisted selection in the peanut breeding program.

2.3 Materials and Methods

2.3.1 Plant materials

The peanut Mini-core collection was constructed by a stratification strategy from the US peanut core collection based on morphological traits (Holbrook and Dong, 2005) were used in this research. This Mini-core collection includes 120 accessions evenly distributed in two subspecies (*hypogaea* and *fastigiata*) , containing four botanical varieties (*hypogaea*, *fastigiata*, *peruviana*, *vulgaris*). The other 12 accessions were selected from the gene bank to present two botanical varieties of *aequatoriana*, and *hirsute* (Table 2.S5).

2.3.2 Descriptive sensory profile

After harvest, the raw peanuts were first dried and mechanically shelled at the USDA, ARS National Peanut Laboratory in Dawson, GA. The dried and shelled mature peanuts were shipped to the USDA-ARS Market Quality and Handling Research Unit in Raleigh, NC for analysis. The raw peanut samples were roasted in a forced-air oven (Model LXD, Dispatch Industries, Minneapolis, MN) at 171°C to a final color of L=48 ±1 on the Hunter scale as determined using a colorimeter (Hunter Labs, Reston, VA). Samples were grounded to a paste through a Robot Coupe Blixer 3 commercial food processor (Robot Coupe USA, Inc., Ridgeland, MS). This preparation eliminated any seed to seed variation (Sanders, et al., 1989).

The pastes were presented to an eight members panel trained in the descriptive evaluation of peanut flavor maintained by the USDA-ARS Market Quality and Handling Research Unit in Raleigh, NC. The panel was trained to use the Spectrum Flavor Descriptive Analysis Technique (Meilgaard, et al., 1999). Samples were presented randomly with three-digit numbers. The flavor descriptors previously described were rated on 15 point scales (Johnsen, et al., 1988; Sanders, et al., 1989). The samples were presented blindly in duplicate, and scores were reported as means.

2.3.3 Sugars profile

The pressed paste recovered from the sampling for tocopherols was completed defatted by boiling hexane in a Soxhlet apparatus. The remaining meal was analyzed for sugars as previously described (Pattee, et al., 2000). In brief, approximately 100 mg of defatted peanut meal was extracted with 15 mL of an aqueous solvent mix (chloroform/methanol/water 60/25/15 v/v/v). The samples were vortexed to mix and then sonicated for 15 min to extract. The samples were then centrifuged at 1000 rpm to pellet the remaining solid material. The solvent layer was decanted to a small beaker and evaporated overnight in a vacuum oven fitted with a solvent trap. The dried residue was brought up with 1 ml of water containing 2.3 mM lactose and 1.2 mM cellobiose as internal standards. The solution was diluted 40 fold with water and filtered through a sulfonic acid column (Dionex On Guard I.I.H., Thermo-Dionex, Sunnyvale, CA) to dislodge free amino acids. Solutions were then analyzed using HPLC allocated with an ion-exchange column and a pulsed amperometric detector (PAD). The HPLC was equipped with Dionex BioLC system fitted with a Dionex PA-1 column (250 mm X 4.6 mm i.d.) and an electrochemical detector. The mobile phase was 200 mM sodium hydroxide in water at 1.0 mL/min. The column was heated to 30°C. For quantitation, a standard mix containing the internal standards, myo-inositol, glucose, fructose, sucrose, raffinose, and stachyose was used. All standards were purchased from Sigma Chemical Corp. (St. Louis, MO). All analyses were done in triplicate and reported as means. Response factors based on the internal standards were calculated and used for the quantitation.

2.3.4 Fatty acids profile

Raw oil content: oil percentage was measured by Maran Pulse nuclear magnetic resonance (NMR). The NMR calculated % oil and % H2O in the sample using 5 and 10 g of whole mature seeds for weighing and further analysis for each entry. The formula for oil percentage calculation was as following: [oil % * 100 / (100 – H2O % * 100)].

Fatty acid composition: We used a coffee bean grinder to grind three to five seeds into a fine powder. Then we extract the oil by transferring almost 150 mg of ground powder into a 16 x 100 mm disposable test tube, and 5.0 ml of n-heptane (Fisher Scientific). We added 500 µl of 0.5 N sodium methoxide (NaOCH3) in methanol solution to the test tube, mixed it with the sample, and converted fatty acids to methyl esters. Then we added 7ml distilled water to separate the organic layer from the aqueous layer and seed residue after 2 hours. To do the GC analysis, we transferred the 1.5 ml aliquot of the organic layer containing the methyl esters to a 2.0 ml autosampler. To determine fatty acid composition, an Agilent 7890A gas chromatograph (GC) equipped with a flame ionization detector (FID) and an autosampler was used. In addition, we used a fatty acid methyl ester (FAME) standard mix RM-3 (purchased from Sigma) to establish peak retention times. Agilent Technologies showed that peak separation performed on a DB-225 capillary column (15 m x 0.25 mm i.d. with a 0.25 µm film). We set helium as the carrier gas to a flow rate of ~1.0 ml/min. One µl of sample was injected at a 60:1 split ratio onto the column maintained isothermally at 210°C. The temperature of inlet and detector were 280°C to 300°C. Each sample took 12 min for a total run. The relative peak areas help us identify and determine the fatty acid composition (Wang et al., 2005).

2.3.5 Tocopherols profile

Raw peanut samples were ground to a fine meal by a household grinder (Krups International, Frankfurt am Main, Germany). Approximately 200 g of meal was wrapped in cheesecloth and loaded into the X and the peanut oil was extracted using a hydraulic machine (Model 2622-1, Carver, Inc., Wabash, IN). The tocopherols in the oil were analyze by normal phase HPLC (Hashim et al., 1993). In a word, 200 mg of oil was weighed into a 2 mL autosampler vial and 0.8 mL of hexane containing 2% (v/v) isopropanol was added. The contents of the vial were vortexed

to mix and injected onto the HPLC (Agilent Model 110, Agilent Technologies, Santa Clara, CA). The mobile phase was 2% (v/v) isopropanol in hexane running isocratically at 1.2 mL/min. The column was a Luna Silica held at a temperature of 30°C. Detection was by UV at 295nm. Solutions of authentic tocopherols in hexane were run simultaneously as standards. Standards of α, γ, and δ tocopherols were purchased from Sigma Chemical Corp. (St. Louise, MO) while β-tocopherol from Matreya L.L.C. (State College, PA). Standards were used to determine retention times and as external standards for calculations of tocopherol contents. All analyses were done in triplicate and reported as means.

2.3.6 Genotyping and quality control

The 120 accessions were genotyped using the Illumina Soy SNP50k iSelect BeadChip (Illumina, San Diego, CA. USA) with a total of 17,723 SNPs as described in Table 2.1. Genotyping Module v1.8.4 (Illumina, Inc. San Diego, CA) were used for SNP alleles calling. Those SNPs with unknown physical location were removed from next-step study. To ensure accuracy of the results, especially for the GWAS analysis below, further data filtering was conducted by eliminating heterozygote SNPs (with missing rates larger than 98%,), SNPs with missing rates larger than 10%, minor allele frequencies less than 5%, as well as SNPs existing in minor states resulting in only two alleles were separating for each SNP site. After the filtering, a total of 11,173 SNPs was kept for further analysis.

2.3.7 SNP calling

This study used Seqtk to process and filter all reads for the purpose of quality control. Burrows-Wheeler Aligner (BWA) was adopted to map the high-quality paired-end reads using the two ancestral species reference genome (A. *ipaensis* and A. *duranensis*) (Li and Durbin, 2009) as well as cultivated peanut reference genome (Zhuang et al., 2019). Realigner Target Creator and InDel-

Realigner in GATK were utilized for InDels realignment, Unified Genotyper for genotypes calling among the 120 accessions with default parameters, the "Depth of Coverage" module of GATK for the sequencing depths calculation for each sample (McKenna et al., 2010), and GATK (McKenna et al., 2010) and SAM tools for single SNPs confirmation (Li et al., 2009). As the peanut genome has allotetraploid characteristics, the genotyping errors caused by partial homologous alignment were adjusted by the first sequencing depth comparison, then filtered those SNPs with genomic integrity (genotyping rate) and minor allele frequency (MAF). This study excluded abnormally high homologous locations where not many SNPs existed.

2.3.8 Statistical analysis and Manhattan plot

In this study, SNP & Variation Suite (SVS, Version 8.0), as well as Genome Association and Prediction Integrated Tool (GAPIT), were used to exploit potential genetic associations between SNP markers and four main sensory and quality attributes. Linkage disequilibrium (LD) pruning strategy was adopted to generate a set of independent SNPs within 50 SNPs (the number of SNPs for each round of LD pruning), a step of five SNPs, with r2 threshold of 0.5 using composite haplotype method (CHM) (Wang et al., 2009). Eleven thousand and one hundred seventy-three (11,173) independent SNPs and LD blocks were screened out after the LD pruning process. Identical by state (IBS) analysis among all pairs of samples were conducted using the independent SNP markers. Principal component analysis (PCA) was used to assess population structure genetic diversity (Price et al.,2006). Then the Efficient Mixed-Model Association eXpedited (EMMAX) algorithm was utilized to examine the sample population structure in the association test with the first two principal components (PC1, PC2). The model is $Y = X\beta- + Z\mu + \varepsilon$, where Y was the vector of sensory attributes, X was the matrix of fixed effects containing the first two PCs, β was the coefficient vector; Z was the matrix of random additive genetic effects, μ was the vector

representing the coefficients of the random effect The other model is Var(μ) = Gσg2, where G was the simple IBS allele-sharing matrix, and σg2 was the additive genetic variance, and ε was the vector of random residuals.

The Manhattan plot was mapped out based on the channel peanut genome sequence (version Coco1.0). This study adopted the Bonferroni-correction to determine the threshold P-value for genome-wide significance (Duggal et al., 2008). Based on that, The LD block was defined as a set of contiguous SNPs with the minimum pairwise r2 value exceeding 0.50 (Gu et al., 2011). After LD pruning, 1,024 independent SNPs and LD blocks were kept. Thus, after Bonferroni correction, the threshold P-value for genome-wide significance was 0.05/1024 = 4.48e-5 (-log10 (P-value) = 4.31), while he threshold P-value for the "suggestive association" was 1/1024 = 9.77e-4(-log10 (P-value) = 3.01), which allowed one false-positive effect in a genome-wide association test.

2.3.9 QTLs and candidate gene search

In order to find the candidate genes, this study examined genes located within 0.5 Mb of the most significant SNPs associated with sensory attributes and identified potential candidates by referencing the peanut genome from PeanutBase (https://peanutbase.org/). The identified genes were annotated using BLAST from the non-redundant protein database (Pruitt et al., 2005).

2.4 Results

2.4.1 Phenotypic analysis among peanut accessions

The phenotypic evaluation displayed a wide range of the natural variations in the 120 peanut accessions from the Mini-core collection (Fig 2.3, Table 2.S5). Four flavors and nutrient-related traits were evaluated in cultivated peanuts, including sensory attributes, sugars, fatty acids, and tocopherols. For most traits, the phenotypic data displayed a near-normal distribution shape. For example, the trait Bitter, ranges from 1.92 to 4.12, with a mean at 2.50; Sweet, varies from 2.06 to

3.64, with an average of 2.79 (Fig 2.3, Fig 2.4). However, some traits, especially in sensory attributes, skewed to the extreme value, like Raw Beany, Ashy, Cardboardy (Table 2.S6). Thus, after the descriptive statistics analysis of phenotypic variation on nine traits listed in Fig 2.3 & Fig 2.4, 23 traits were selected to conduct further analysis, which abided by a normal distribution and with enough data (missing data less than 10%). These 23 traits encompassed six categories from sensory attributes profile, six from sugars profile, six from fatty acids profile, and five traits from the tocopherols profile.

2.4.2 Peanut genomes and SNP profiles

A total of 18,426 high-quality SNP markers uniformly spreading on A sub-genome and B sub-genome were acquired from the 120 accessions. After filtering out the SNPs located on scaffolds, 17,232 high quality SNPs with MAF > 5% and integrity > 50% were chosen for next-step analyses (Table 2.1). The filtered SNP markers, though, were not homogeneously distributed across the whole peanut genome, with 7,516 and 8,914 SNPs on A and B sub-genomes, respectively. Chromosome B09 possessed the highest amount of SNPs (9.62%; 1,658 of 17,232), while shortest chromosome A08, with a genome size of 51.90 Mb, contained the least (2.12%; 365 of 17,232). The average numbers of SNPs/Mb were seven and eight on A and B sub-genomes, respectively. While the average genes/Mb were both 35 for two sub-genomes. In detail, Chromosome B03 had the highest number of genes (5,188 of 77,617), followed by its counterpart, chromosome A03, with the second-highest number of genes (4,929 of 77,617) (Table 2.1).

A total of 310, 474, 569, and 230 associated QTLs were identified through GWAS analysis with sensory attributes profile, sugars profile, fatty acids profile, and tocopherols profile, respectively, including 14, 87, 105 and 73 significant QTLs for each profile (Table 2,2). Among the significant

QTLs, there were 4, 55, 86, and 49 QTLs with PVE higher than 15% for each profile (Table 2.2), indicating a higher heritability for these traits.

2.4.3 US Mini-core collection and population structure

Unlike many other crops, the cultivated peanut is of hybrid origin and has an allotetraploid genome that comprises two complete sets of chromosomes from two ancestral wild diploid species, A. *duranensis* (A-genome) and A. *ipaensis* (B-genome) (Bertioli et al., 2019). Among 120 genotypes, there were six botanical varieties, *fastigiata*, *hypogaea*, *peruviana*, *vulgaris*, *aequatoriana*, and *hirsute* (Table 2.S5). Despite some mall divergence, the population structure is obviously related to the botanical variety. Based on ΔK information from STRUCTURE analysis, we chose K = 6 as optimal stratification (Fig. 2.1.A). After 10 runs with K = 6, the run with the highest likelihood value was chosen to allocate the following membership coefficients (Q) to each genotype, and as a result, a graphical bar plot was generated with Q with six subpopulations, named with G1, G2, G3, G4, G5 and G6 (Fig. 2.1.B).

2.4.4 Sensory attribute profile

Nine flavor related traits were chosen for the identification of potential genetic loci and regions, while only six traits were finally kept, because of their values abided by a normal distribution, which were Sweet, Bitter, Sweet Aromatic (SA), Raw Beany (RB), and Woody Hulls Skins (WHS). In six sensory attribute traits, significant QTLs were only found in sweet and bitter traits by the Manhattan plots based on EMMAX method (see Fig 2.6) and quantile-quantile plots (see Fig 2.5), in which the13 QTLs on linkage groups A03, A04, A09, B04, and B06 significantly associated with trait Sweet and one QTL on linkage group A03 with trait Bitter (- log10 (P-value) > 4.31). However, according to the Bonferroni correction (-log10 (P-value) = 3.01), the 310 SNP associated with six flavor traits achieved the corrected P-value (Fig 2.6, Table 2.2 & 2.3). A 0.5 Mb genomic

region on both sides of the peak SNPs, which associated with the corresponding traits, was subsequently analyzed for the exploration of potential genes (Table 2.8). The distribution of all 310 QTLs across 20 linkage groups (LG) revealed that 171 QTLs were related to Sweet trait, while 139 QTLs were related to five other traits (Table 2.S1). Especially, 13 of the 14 significant QTLs were found significantly associated with the Sweet trait, whereas one was found significantly associated with the Bitter trait. No more significant associations were found between QTLs and the other four traits. This result indicated that the Sweet trait has the highest heritability comparing with other sensory attributes, which is consistent with past papers saying roasted peanut sweet attribute was a highly heritable trait (Pattee et al., 1998). In the meanwhile, 59 and 48 QTLs were suggestively mapped throughout LGA04 and LGB04, respectively, indicating that these two chromosomes have more related genomic regions than the other chromosomes. Also, one genomic region, AX-177643393 on LGA03, was found significantly related to both Sweet and Bitter traits. A total of 101 genes were identified associated with sensory attributes, including 13 genes in significant genomic regions, and 85 in suggestive regions. A total of 13 genes were identified functionally associated with sensory attribute within the 14 significant genomic regions, where six genes located in LGA04 and four genes located in LGB03. Of the 13 genes, ten genes were functionally associated with trait Sweet, three genes associated with trait Bitter. All these genes were annotated to involve in biological macromolecule metabolic processes, such as carbohydrate and phenols.

For other sensory attributes such as SA, RB, and WHS, there were 92 SNPs identified to reach the suggestive value, with 63 QTLs associated with SA, 24 QTLs associated with RB, and 5 QTLs with WHS, although they did not reach the significant corrected P-value. Besides, 43 genes were found associated with related traits in these suggestive regions, with 16 genes functionally

associated with SA, 19 genes associated with RB, and 8 genes associated with WHS. Besides, 15 genes were also screened within the suggestive region for trait sweet, as well as 24 genes for trait bitter (Table 2.8). Interestingly, previous studies showed significant correlations among sweet, bitter, and roasted peanut attributes as well. It was reported that there was a positive correlation between total sugars and sweetness, while a negative correlation between total sugars and bitterness, which is consistent with this study (Pattee et al., 2000).

2.4.5 Sugars Profile

The Manhattan plots drawn by EMMAX method (See Fig 2.7) and the Q-Q plots (See Fig 2.5) demonstrated that there were 36 QTLs significantly associated with trait Sucrose, 22 QTLs associated with Fructose, 15 QTLs associated with Raffinose, and 14 QTLs associated with Total Sugar ((-log10 (P-value) > 4.31). The distribution of all 79 QTLs across 17 linkage groups (LG) revealed that 34 QTLs were distributed throughout 9 LGs of the A sub-genome. The rest 45 QTLs were mapped across 8 LGs of the B sub-genome (Table 2.2& 2.3), indicating both A sub-genome B sub-genome plays an essential role in genomic regions that are related to the carbohydrate biosynthesis process.

Of 36 Sucrose related QTLs, six were found in LGA04, nine were found in Chromosome B04 and, five were found in LGB05. As shown in Table 2.5, their PVE ranged from 14.37% to 22.56%, while MAF ranged from 0.06 to 0.13. However, the highest SNP AX-147233387 (- log10 (P-value = 9.08)) was located at position 57,329,521bp on LGA09, with PVE 29.57 and minor allele frequencies 0.06. Except for LGA04, LGA09, LGB04, and LGB05, significant genomic regions on the other ten linkage groups were also found associated with Sucrose by EMMAX method, with one to two QTLs on each linkage group. Interestingly, like sucrose, Fructose, and Total Sugar have similar significant QTLs distribution patterns identified in linkage groups. There were three

QTLs in LGA04 and five in LGB04 of 14 significant QTLs for trait Fructose, while four QTLs in LGA04 and four in LGB04 of 14 significant QTLs for trait Total Sugar. However, trait Raffinose had unique significant QTLs distribution patterns identified in linkage groups, with eight significant QTLs in LGB09 of 15 in total (Table 2.5).

One hundred and thirty-nine genomic regions on different linkage groups were found to be suggestively associated with trait Sucrose ((− log10 (P-value) > 3.01), but there were no statistically significant at the genome level. As shown in Table 2.S2, LGA04 harbored more suggestive QTLs (40) than other linkage groups in a long-range genomic region from 27,200,581 bp to 120,352,138 bp, with PVE ranging from 9.62 to 14.16. Followed by LGB04, LGB05 contained 29 and 20 QTLs reaching suggestive genome-wide significance, with PVE ranging from 9.98 to 14.10, and from 9.67 to 14.16, respectively. Since the peanut genome sequence is available, genes were determined within the region containing both significant and suggestive QTLs. Similarly, 89, 49, and 56 genomic regions on different linkage groups were found to be suggestively associated with trait Fructose, Raffinose, and Total Sugar ((− log10 (P-value) > 3.01). There were also 62 genomic regions found to be suggestively associated with trait Inositol, shown in Table 2.9.

Within 1Mb, 21 genes were found close to the QTLs, which were significantly associated with Sucrose of the 37 genetic regions (Table 2.S2). Of the 21 genes, 17 were found to have known functions in carbohydrate metabolic processes, such as B363MR (sucrose phosphate synthase), THTJ31 (NAD-dependent malic enzyme), and HLUA2X (Malic oxidoreductase). There was also one gene found to have functions in the 070EBZ (SWEET sugar transporter). Similarly, 8, 8, and 21 genes were found nearby significant QTLs associated with trait Fructose, Raffinose, and Total Sugar, respectively (Table 2.9). Sucrose-phosphate synthase (SPS) was reported as a key

regulatory enzyme in the pathway of sucrose biosynthesis and is also linked to quantitative trait loci controlling plant growth and yield, suggesting biological multifunction for this gene (Castleden et al., 2004). Besides, Lin et al. reported that SWEET9, belongs to SWEET sugar transporter, could function in (1) sucrose efflux from the phloem, (2) sugar uptake into, and (3) sugar efflux from nectary parenchyma.

2.4.6 Fatty Acids Profile

According to the result of GWAS analysis, a total of 569 QTLs were identified related to Fatty Acids Profile, of which 105 are significant QTLs (Table 2.6, Fig 2.5 & Fig 2.8). The distribution of all 569 QTLs across 20 linkage groups revealed that 194 QTLs were distributed through 10 LGs of the A sub-genome, and 375 QTLs were mapped across 10 LGs of the B sub-genome (Table 2.S3). This indicated that the B sub-genome has more genomic regions associated with the fatty acid biosynthesis process than the A sub-genome.

There were 31 significant QTLs dispersed on nine linkage groups associated with trait IV value ((% oleic * 0.8601) + (% linoleic * 1.7321 + % eicosenoic * 0.7854)) (Hashim et al., 1993). 14 of them were concentrated on LGB10 in a genomic region from 111,329,372 to 126,438,874bp, spanning approximately 15.11Mb with PVE ranging from 14.32% to 18.73%, MAF ranging from 0.16 to 0.35 and − log10 (P-value) ranging from 4.43 to 5.63 (Table 2.3). The same QTLs distribution pattern appeared for trait O/L ratio (% oleic (18:1)/ linoleic (18:2)), with nine QTLs concentrated on LGB10 spanning approximately 12.42 Mb. As shown in Table X, their PVE ranged from 14.45% to 15.33%, and MAF ranged from 0.16 to 0.35. The explanation of this phenomenon could be that both IV value and O/L ratio were calculated based on the content of oleic acid and linoleic acid. The results in which the QTLs were found associated with trait P/S ratio (% palmitic (16:0)/ stearic (18:0)), LGA03 (8), LGB03 (22), and LGB10 (13) harbored 43 of

the 50 significant QTLs in total, indicated these three linkage groups play an important role in genomic regions in the metabolic regulation process between palmitic acid and stearic acid. Two genomic regions (AX- 176815995and AX- 176822343) on LGB05 were found to be significantly related to trait Saturation rate (% saturated FA/ total FA), and they explained 15.90% and 15.24% of the phenotypic variance. One genomic region (AX- 147263936) located on LGB03 and explained 14.83% of the phenotypic variance.

To explore the potential genes involved in Fatty Acids profile, this study examined the ± 0.5 Mb regions, which were around the SNPs nearby significant QTLs. A total of 35 genes nearby significant QTLs associated with Fatty Acids profile were determined, including 16 genes associated with trait IV value, six genes associated with O/L ratio, 11 genes associated with P/S ratio, and two genes associated with the Saturation rate (Table 2.10). Among these genes, 23 genes were found to have known functions in fatty acid biosynthetic process and lipid metabolic process. Interestingly, all six genes found associated with O/L ratio were also identified in the genes associated with IV value, including X66CUM (3-ketoacyl-CoA synthase 10), 5IXZ6F (sphingolipid delta desaturase), MF24RD (enoyl-CoA hydratase/isomerase family protein), 133ZSG (3-ketoacyl-CoA synthase), V2BP3N (3-ketoacyl-CoA synthase), FLDD49 (3-hydroxyisobutyryl-CoA hydrolase-like protein 5-like isoform X2). Two genes (133ZSG and V2BP3N) were found to be associated with the Saturation rate, and one gene (3ZW7IP) was found to be associated with P/S ratio were also found in the genes associated with IV value (Table 2.10).

2.4.7 Tocopherols Profile

Tocopherols are a series of naturally occurring compounds that serve to provide antioxidant protection to the fatty acids present in peanuts. According to the Manhattan plots (Fig 2.9) and Q-Q plots (Fig 2.5), 70 significant QTLs were found to be associated with the Tocopherols profile.

Same as the sugars profile, the distribution of all 70 QTLs across 19 linkage groups (LG) suggested that 32 QTLs were distributed throughout 9 LGs of the A sub-genome. 38 QTLs were mapped across 10 LGs of the B sub-genome (Table 2.7), indicating both A sub-genome B sub-genome plays a crucial role in genomic regions in the tocopherol biosynthesis process. Besides, 157 suggestive QTLs were also identified to be associated with the Tocopherols profile. On LGB06, there were five significant QTLs associated with α-Tocopherol in a genomic region from 104,214,231 to 117,556,196 bp, spanning approximately 13.35 Mb, as shown in Table 3, with PVE ranged from 17.38% to 17.60% and − log10 (P-value) ranged from 5.23 to 5.29. Followed by LGA06, 3 QTLs reached genome-wide significance, with PVE ranged from 15.83% to 17.60%, and minor allele frequencies (MAF) ranged from 0.05 to 0.06. Except for LGB06 and LGA06, other SPNs close to significant QTLs were dispersed on LGA03 (1), LGA04 (1), LGA08 (1), and LGB01 (1). A total of 30 SNPs nearby significant QTLs were found to be associated with β-Tocopherol, of which seven QTLs concentrated on LGA03 and six located on LGB05, with PVE ranged from 14.24% to 14.80%, and from 16.90% to 18.13% respectively. The other significant QTLs were distributed on different linkage groups, and the ratio of phenotypic variation was explained by the QTLs varied from 14.17% to 22.41%. Two genomic regions (AX-147241099 and AX- 176791511) on LGB02 were found to be significantly related to δ -Tocopherol, and they explained both 15.48% of the phenotypic variance. LGA09 and LGB09 harbored three and four significant QTLs of the 13 in total related to γ-Tocopherol, respectively. However, the most significant SNP (AX- 176805798) located in LGB06, with MAF 0.06 and PVE 27.56. One genome-wide significant region for Total Tocopherols was detected on LGB06. The genome-wide significant region harbored five QTLs that were statistically significant at the genome level (− log10 (P-value) > 4.31). The significant SNPs were in a genomic region from 104,214,231 to

117,556,196 bp, spanning a total of approximately 13.35 Mb with PVE ranged from 18.59 to 19.10% (Table 2.7). Besides, other SPNs close to significant QTLs associated with Total Tocopherols were dispersed on LGA04 (2), LGA06 (1), LGA08 (1), LGB05 (1) and LGB09 (1), which is similar to linkage groups distribution pattern for α-Tocopherol, indicating α-Tocopherol accounts for a large proportion of Total Tocopherols.

1 Mb regions around SNPs nearby significant QTLs were also examined to screen the potential genes involved in the Tocopherols profile. A total of 23 genes was identified to be significantly associated with Tocopherols, including five genes functionally associated with α-Tocopherol, eight genes associated with β-Tocopherol, two genes associated with δ -Tocopherol, five genes associated with γ -Tocopherol, and three genes associated with Total Tocopherols (Table 2.11).

2.5 Discussion

The min-core of the peanut germplasm collection is a set of samples that serve as a manageable representative selection to represent the overall collection based on morphological traits (Holbrook and Dong, 2005). No previous study has been conducted to investigate the flavor related traits for the Mini-core. The chemistry of peanut flavor has been studied intensely but has not been completely elucidated. Roast flavor in many foods has been attributed mainly to the Maillard reaction, which is the polymerization of free amino acids with sugars under heat. These compounds have been identified in peanuts (Chiou, et al., 1993), but the use of them to produce peanut flavor in model systems was not successful.

With more DNA markers available in peanut, genome-wide association mapping is becoming feasible. SNP coverage on the genome and genotype diversity panel sizes can affect the ability to achieve significance. The SNP numbers used in this study were relatively small. GWAS has a higher power to detect associated markers, but it is also can produce false-positive associations,

which is the type I errors (Campbell et al., 2005; Zhong et al., 2017). EMMAX model, using high-density markers to calculate a pairwise relatedness, was utilized for correcting population stratification, which can lead to biased or spurious results (Kang et al., 2010). Hence, the EMMAX model was used and modified for the first ten PCs after the kinship matrix-pairwise IBS distance calculation, to eliminate the false-positive results caused by the population structure observed in this study. It has been widely applied in association study and yields more accurate statistics than other methods (Zhou and Stephens, 2012). The statistical significance of suggestive QTLs detected in this study may also be affected by phenotyping errors. PCA was used to quantify the population structure of these 120 accessions (Fig 2.S3). Six distinct clusters were presented in the figure, which indicates that these Mini-core collections represent a highly structured population.

In our study, 1583 significant QTLs related to four main sensory and quality attributes were detected, and the used diversity panel is suitable for GWAS.

2.5.1 Sensory Attributes profile

As more powerful strategies being utilized for the discrimination and description of both the qualitative and quantitative sensory attributes in roasted peanuts, such as the descriptive sensory analysis by trained panelists (Meilgaard et al., 1999; Pattee et al., 2001), there are more and more publications characterizing the variation in sensory attributes in recent years. Highly significant correlations have been found among the least square means for the sensory attributes, such as bitter with sweet in roasted peanut (Pattee et al., 1997, 1998). Furthermore, certain roasted peanut sensory and quality attributes have been shown to be heritable traits (Pattee et al., 1995, 1998), with highly significant correlations among least square means for the attributes (Pattee et al., 1997, 1998), with broad-sense heritability in U.S. peanut germplasm sources ranging from 10.6% to 24.3% (Pattee et al., 1995). These previous results inspire strong confidence to breeders the enormous

potential for improving the sensory and quality attribute levels through various proper breeding strategies.

In this study, 19 SNP peaks along with potential genes and QTLs were determined associated with corresponding sensory attribute traits (Fig 2.3 and Table 2.S3). Within these genetic regions, a total of 310 SNPs associated with six traits reached the corrected P-value (-log10 (P-value) = 3.01) (Fig 2.6, Table 2.3), according to the Bonferroni correction. However, only 12 QTLs identified are significantly associated with the sensory attributes, with 11 QTLs associated with the Sweet trait, and one with the Bitter trait. There were no significant QTLs identified associated with the other four traits. This result indicated that the Sweet trait has the highest heritability comparing with other sensory attributes, which is consistent with earlier papers saying roasted peanut sweet attribute had the highly heritable traits (Pattee et al., 1998).

More significant QTLs found on trait Sweet, with 7 SNP localized on LGA03, LGA04, LGA09, and LGB04, indicating there is a high heritability for this trait. There were 48 genes with 1Mb from peak SNP AX-147248776 (Fig 2.10), explaining 17% of the PVE. Within this genomic region, FMT16N (130835613-130841090 bp), known as transferring glycosyl group transferase, was found to locate in Chromosome B04, which catalyzed the transfer of sugar residues from an activated donor substrate to an acceptor molecule (Saxena et al., 1995). There are two more critical genes screened within 1 Mb of the genetic region. For example, the ACJ5YK, known as transferring cinnamyl alcohol dehydrogenase homolog, catalyzed the final step in the biosynthesis of monolignols (Ma, 2010). The 7P4FH0, which helps code probable polygalacturonase, involves in carbohydrate biosynthetic process. Some other essential genes, such as 7P4FH0 (probable polygalacturonase), VJ6UAM (Aldehyde/histidinol dehydrogenase), FDLX2Z (Glycosyltransferase), 42PK2U (Fasciclin-like arabinogalactan family protein), 2X1NQK

(Glycosyltransferase), 3PMI9T (sterol 3-beta-glucosyltransferase), FW69P6 (Glucose/ribitol dehydrogenase), 31KTN8 (Galactose oxidase), were also found participating in carbohydrate metabolism in other significantly SNPs regions. Besides, genomic regions on other linkage groups were also found to be suggestively associated with Sweet by EMMAX method and obtained nine genes distribute in 5 chromosomes (Table 2.8 & 2.S1).

Interestingly, all 11 significant QTLs associated with trait Sweet, including AX-147248776, were also identified significantly with trait Sucrose and Total Sugar, with PVE ranging from 14.66% to 29.93%, and 14.98 to 21.49%, respectively, indicating the sucrose and total sugars play fundamental roles in Sweet trait. Another interesting phenomenon found was that there was an appreciable quantity of QTLs related to disease located around the peak SNPs. Around SNP peak SNP B4-AX-147248776, there were 23 genes related to disease resistant, which including Y7I7Z8, IB6E9K, BW1AWE, 27QVSF, IZ5HKP, Q1WEA7, EG26MQ, 59G75N, Q05SK2, QZS0X1, RGV79Z, 27BWBV, N2E3I0, UZQ44S, UAH69W, F5ZG0R, XA8QRS, DKE03A, occupying 54% of total genes. In order to identify potential QTLs for thrips, TSWV and leaf spot in peanut, Wang et al. in 2013 constructed two genetic maps, in which, one QTL for TSWV, qF2TSWV3, was found in the same genetic region (seq5D5-GM2744) on linkage group AhII with another QTL for leaf spot, qF2LS1. Besides, another QTL, qF5LS10 for leaf spot reported in this article, was located between GM1254 and seq15C10 on linkage group LGT17. This finding shows that this genetic region should involve in disease resistance, indicating there is specific functional internal relation between carbohydrate metabolic process and disease resistance.

For the Bitter trait, a total of 17 QTLs were discovered with one significant QTLs and 16 suggestive QTLs distributed on different linkage groups (Table 2.4). The significant SNP AX-176814258 was also associated with trait Sweet, indicating these two traits share a linked pathway

for related component metabolic process. The SNP located at 12720888 bp in LGA03, with a minor allele frequency of 0.06 and explaining 15.68% of the phenotypic variance. Other suggestive peak SNPs were distributed on LGA01, LGA06, LGA07, LGA08, LGA10, LGB03, LGB06, LGB07, LGB08, and LGB10. The candidate genes in the peanut genome sequence of 1Mb windows (SNP position ± 0.5 Mb) and surrounding each identified promising SNP were obtained. A total of 27 genes were determined, which included three genes nearby significant QTLs and 24 genes around suggestive QTLs. The corresponding genomic positions and biological processes related to the flavor genes are listed in Table S3 and Table S6. Within the significantly genetic region, three genes, 2X1NQK (Glycosyltransferase), EZ9WFK (strictosidine synthase activity), G490LK (arabinogalactan protein), have the known functions in carbohydrate and related compound metabolism. Besides, 24 functionally related genes were found around the suggestive peak SNPs, including several important genes involved in sugar synthesis, such as CKCA5J (Glucose-1-phosphate adenylyltransferase, reported function in glycogen biosynthetic process), G1I4C8 (located in LGB8, reported function in seed storage protein and nutrient reservoir activity). Some genes within 1MB of the significant QTLs were reported in other agricultural traits regulation pathways, like disease resistance, suggesting that the regulation of these compound synthetic is involved in the disease defense pathway.

2.5.2 Sugars Profile

The sugars (simple carbohydrates) in peanuts serve to provide energy for the growing seed. As a result, maturity is the major effector for the concentration of sugars in peanut. The maturity of a peanut crop at harvest is under the influence of the genetics of grown cultivar and the environmental conditions of growth. Sugar levels have been closely related to positive flavor descriptors such as sweet and sweet aromatic as well as to negative descriptors such as fruity

fermented (Pattee et al., 2000; Greene et al., 2008). Interestingly, we found that all 11 significant QTLs associated with trait Sweet was also identified significantly with trait Sucrose and Total Sugar, with PVE ranging from 14.66% to 29.93%, and 14.98 to 21.49%, respectively.

A total of 87 QTLs were identified significantly associated with Sugars profile, including 36 QTLs significantly associated with trait Sucrose, 22 QTLs with Fructose, 15 QTLs with Raffinose, and 14 QTLs with Total Sugar ((− log10 (P-value) > 4.31). Moreover, all the 17 significant QTLs associated with trait Total sugar were identified in trait Sucrose, indicating that the sucrose biosynthesis process has more significant in the whole sugar biosynthesis process than raffinose and other sugars. Twenty-one genes were found close to the significant QTLs associated with Sucrose (Table 2.S2). Among these 21 genes, 17 were found to have known functions in carbohydrate metabolic process, including B363MR (sucrose phosphate synthase), THTJ31(NAD-dependent malic enzyme), and HLUA2X (Malic oxidoreductase). Specifically, there are also three genes involved in glucose metabolic process, including 74FNJK (glucose-6-phosphate dehydrogenase), PV44JT (glyceraldehyde-3-phosphate dehydrogenase) and 9E7XBZ (glucose-6-phosphate dehydrogenase), two genes involved in mannose metabolic process, including RSDZ0N (Galactose mutarotase-like domain) and 0R79PA (NAD(P)-binding Rossmann-fold superfamily protein). Besides, an important gene, 070EBZ, was screened to have functions in the SWEET sugar transporter, which plays a vital role in sugar transport in a plant cell. Besides trait Sucrose, 8 and 21 genes were found nearby significant QTLs associated with trait Fructose, Raffinose, respectively, including 5QB4T4 (sugar transporter), function in transmembrane transporter activity (Table 2.9).

2.5.3 Fatty acids profile

569 QTLs are reaching the suggestive line (-log10 (P-value) = 3.01), with 194 QTLs distributed through 10 LGs of the A sub-genome, and 375 QTLs mapped across 10 LGs of the B sub-genome (Table 2.S3). Like suggestive QTLs, among 104 significant QTLs 75 located on the B sub-genome and 29 located on the A sub-genome (Table 2.6). This indicated that the B sub-genome has more genomic regions associated with the fatty acid biosynthesis process than the A sub-genome. Moreover, the significant QTLs concentrate upon LGB03 and B10, possessing 31 and 38 QTLs, respectively. Precisely, 15 of the 31 significant QTLs associated with IV were mapped on LGB10, while nine of the 20 significant QTLs associated with O/L ratio were on LGB10, as shown in Table 3, suggesting the LGB10 participates in regulating the mutual transformation of oleic acid and linoleic acid. For trait, P/S ratio, LGB03 (22), and LGB10 (13) harbored 35 of the 50 significant QTLs, indicated these two linkage groups play indispensable roles in genomic regions in the metabolic regulation process between palmitic acid and stearic acid.

Among the genes identified within ± 0.5 Mb regions of the significant QTLs, 23 genes were found to have known functions in fatty acid biosynthetic process and lipid metabolic process. Interestingly, all six genes found associated with O/L ratio were also identified in the genes associated with IV value, including X66CUM (3-ketoacyl-CoA synthase 10), 5IXZ6F (sphingolipid delta desaturase), MF24RD (enoyl-CoA hydratase/isomerase family protein), 133ZSG (3-ketoacyl-CoA synthase), V2BP3N (3-ketoacyl-CoA synthase), FLDD49 (3-hydroxyisobutyryl-CoA hydrolase-like protein 5-like isoform X2). Two genes (133ZSG and V2BP3N) were found to be associated with the Saturation rate, and one gene (3ZW7IP) was found to be associated with P/S ratio were also found in the genes associated with IV value (Table 2.10).

2.5.4 Tocopherols profile

Tocopherols are a series of naturally occurring compounds that serve to provide antioxidant protection to the fatty acids present in peanuts. Additionally, especially α-tocopherols, have vitamin E bioactivity for humans. Peanuts contain substantial levels of tocopherols that approach those of tree nuts and have higher levels of γ-tocopherol (Maguire et al., 2004). The γ-form provides the highest levels of antioxidant activity to plant oils (Cooney et al., 1993; Jonnala et al., 2006). Although there may be some genetic forces in effect, the levels of tocopherols have been reported to be environmentally influenced (Baydar et al., 2005; Jonnala et al., 2006). Besides, tocopherols could be only produced by photosynthetic organisms, containing all plants, algae, and most cyanobacteria (Horvath et al., 2006).

Previous studies showed that there are two main tocopherol biosynthetic pathways in plant: Shikimate pathway, using Chorismate and Tyrosine as the precursor substances, and methylerythritol phosphate (MEP) pathway, using Pyruvate, Glyceraldehyde 3-phosphate, geranylgeraniol, and Chlorophyll A as the precursor substances (Qi et al., 2005, Valentin et al., 2006). Another tocopherol biosynthetic pathway was also reported in a plant, which is mediated by Cytochrome P450 ω-Hydroxylase (Sontag et al., 2007; Bardowell et al., 2012). In our study, a total of 23 genes was identified to be significantly associated with Tocopherols, including five genes functionally associated with α-Tocopherol, eight genes associated with β-Tocopherol, two genes associated with δ -Tocopherol, five genes associated with γ -Tocopherol, and 3 genes associated with Total Tocopherols (Table 2.11). Among these genes identified, four genes were reported involved in Shikimate pathway, including QYLY8C (Tyrosine transaminase family), DLM7XC (Tyrosine aminotransferase), 36ZVTJ (Prephenate dehydratase) and 71EM27 (Prephenate dehydratase), seven genes were reported involved in MEP pathway, including 308EKY (phosphoenolpyruvate carboxylase), A8AWVV (pyridoxal kinase), QYLY8C (Pyridoxal

phosphate-dependent transferase), FXLQ18 (pyruvate decarboxylase), 79DDAD (peptidyl-prolyl cis-trans isomerase CYP40-like isoform X1), 989M5Y (pyruvate orthophosphate dikinase), and WLW17E (E2 component of pyruvate dehydrogenase), and the genes were reported involved in Cytochrome P450 ω-Hydroxylase mediated pathway, including ZZ7V32 (Cytochrome P450 superfamily protein), 3GHL6G (Cytochrome P450 superfamily protein), M1JRW6 (Cytochrome P450 superfamily protein), 9P7KAZ (Cytochrome P450 superfamily), E313MA (Cytochrome P450 superfamily protein), U5WKYJ (Cytochrome P450) and W3WWNZ (Cytochrome P450 superfamily). IS0G8W, suggestively associated with three SNP makers (AX-147233960, AX-176797662, AX-147233770), encodes a key enzyme, phytol kinase, in MPE pathway, which catalyzes chlorophyll phytol to phytyl diphosphate. A recent study shows that down-regulation of tomato PHYTOL KINASE (VTE5) strongly impairs tocopherol biosynthesis and affects prenyl lipid metabolism in an organ-specific manner (Almeida et al., 2016). Besides, VTE5 deficiency was significantly impacted lipid metabolism, including prenylquinones, carotenoids, and fatty acid phytyl esters (Edreva et al., 2005; Demmig-Adams et al., 2014).

However, although high-quality genome assembly for peanut (*Arachis hypogaea*), cultivar 'Tifrunner,' and the corresponding genome annotation were finished in recent (Bertioli et al., 2019), the annotation is still incomplete. Hence, some genes of critical enzymes involved in tocopherol biosynthetic pathway could even not be screened in peanut genome annotation, such as dimethylallyl diphosphate, isopentenyl diphosphate, methylerythritol phosphate, tocopherol cyclase, homogentisate phytyltransferase, and 2-methyl6-phytylbenzoquinol methyltransferase (Qi et al., 2005; Valentin et al., 2006).

2.6 References

Almeida, J., Da Silva Azevedo, M., Spicher, L., Glauser, G., Vom Dorp, K., Guyer, L., … Rossi, M. (2016). Down-regulation of tomato PHYTOL KINASE strongly impairs tocopherol biosynthesis and affects prenyllipid metabolism in an organ-specific manner. Journal of Experimental Botany, 67(3), 919–934. https://doi.org/10.1093/jxb/erv504

Bardowell, S. A., Ding, X., & Parker, R. S. (2012). Disruption of P450-mediated vitamin E hydroxylase activities alters vitamin E status in tocopherol supplemented mice and reveals extra-hepatic vitamin E metabolism. Journal of Lipid Research, 53(12), 2667–2676. https://doi.org/10.1194/jlr.M030734

Baydar, H., & Erbaş, S. (2005). Influence of seed development and seed position on oil, fatty acids and total tocopherol contents in sunflower (Helianthus annuus L.). Turkish Journal of Agriculture and Forestry, 29(3), 179–186. https://doi.org/10.3906/tar-0406-3

Belamkar, V., Selvaraj, M. G., Ayers, J. L., Payton, P. R., Puppala, N., & Burow, M. D. (2011). A first insight into population structure and linkage disequilibrium in the US peanut minicore collection. Genetica, 139(4), 411–429. https://doi.org/10.1007/s10709-011-9556-2

Bertioli, D. J., Jenkins, J., Clevenger, J., Dudchenko, O., Gao, D., Seijo, G., … Pandey, M. K. (2019). The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. Nature Genetics, 51(5), 877.

Bertioli, D. J., Ozias-Akins, P., Chu, Y., Dantas, K. M., Santos, S. P., Gouvea, E., *et al.,*(2014). The use of SNP markers for linkage mapping in diploid and tetraploid peanuts. Genes Genome Genet. 4, 89–96. doi: 10.1534/g3.113.007617

Botstein, D., White, R. L., Skolnick, M., & Davis, R. W. (1980). Botstein. Am J Hum Gen, 32, 314–331. Retrieved from papers2://publication/uuid/0B80518E-A22B-41F3-BE43-171F51007E42

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics, 23(19), 2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23, 2633–2635. doi: 10.1093/bioinformatics/btm308

Chang, P. K., Wilson, R. A., Keller, N. P., & Cleveland, T. E. (2004). Deletion of the Δ12-oleic acid desaturase gene of a nonaflatoxigenic Aspergillus parasiticus field isolate affects conidiation and sclerotial development. Journal of Applied Microbiology, 97(6), 1178–1184. https://doi.org/10.1111/j.1365-2672.2004.02424.x

Chiou, R.Y.Y., Cheng, S.L., Tseng, C.Y., and Lin, T.C. 1993. Flavor fortification and characterization of raw peanut oils subjected to roasting with partially defatted peanut meal under various atmospheric conditions. J. Agric. Food Chem. 41(7), 1110-1113.

Collard, B. C. Y., & Mackill, D. J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences, 363(1491), 557–572. https://doi.org/10.1098/rstb.2007.2170

Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica, 142(1–2), 169–196. https://doi.org/10.1007/s10681-005-1681-5

Cooney, R. V., Franke, A. A., Harwood, P. J., Hatch-Pigott, V., Custer, L. J., & Mordan, L. J. (1993). γ-Tocopherol detoxification of nitrogen dioxide: Superiority to α- tocopherol.

Proceedings of the National Academy of Sciences of the United States of America, 90(5), 1771–1775. https://doi.org/10.1073/pnas.90.5.1771

Demmig-Adams, B., Cohu, C. M., Amiard, V., van Zadelhoff, G., Veldink, G. A., Muller, O., & Adams, W. W. (2013). Emerging trade-offs - impact of photoprotectants (PsbS, xanthophylls, and vitamin E) on oxylipins as regulators of development and defense. New Phytologist, 197(3), 720–729. https://doi.org/10.1111/nph.12100

Duggal, P., Gillanders, E. M., Holmes, T. N., & Bailey-Wilson, J. E. (2008). Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. BMC Genomics, 9, 1–8. https://doi.org/10.1186/1471-2164-9-516

Edreva, A. (2005). Generation and scavenging of reactive oxygen species in chloroplasts: A submolecular approach. Agriculture, Ecosystems and Environment, 106(2-3 SPEC. ISS.), 119–133. https://doi.org/10.1016/j.agee.2004.10.022

Elements, S., Capabilities, D., & Leadership, D. (2006). Building digital capability : Being an effective digital leader. 18(January), 212–224. https://doi.org/10.1105/tpc.105.037077.beneficial

Greene, J. L., Sanders, T. H., & Drake, M. A. (2008). Characterization of volatile compounds contributing to naturally occurring fruity fermented flavor in peanuts. Journal of Agricultural and Food Chemistry, 56(17), 8096–8102. https://doi.org/10.1021/jf800450k

Gu, J., Chen, M., Shete, S., Amos, C. I., Kamat, A., Ye, Y., … Wu, X. (2011). A genome-wide association study identifies a locus on chromosome 14q21 as a predictor of leukocyte telomere length and as a marker of susceptibility for bladder cancer. Cancer Prevention Research, 4(4), 514–521. https://doi.org/10.1158/1940-6207.CAPR-11-0063

Hashim, I. B., Koehler, P. E., & Eitenmiller, R. R. (1993). Tocopherols in runner and virginia peanut cultivars at various maturity stages. Journal of the American Oil Chemists' Society, 70(6), 633–635. https://doi.org/10.1007/BF02545333

Holbrook, C.C. and Dong, W. 2005. Development of a Mini-core collection for the U.S. peanut germplasm collection. Crop Sci. 45, 1540-1544.

Horvath, G., Wessjohann, L., Bigirimana, J., Jansen, M., Guisez, Y., Caubergs, R., & Horemans, N. (2006). Differential distribution of tocopherols and tocotrienols in photosynthetic and non-photosynthetic tissues. Phytochemistry, 67(12), 1185–1195. https://doi.org/10.1016/j.phytochem.2006.04.004

Hudson, R. R., Slatkin, M., and Maddison, W. P. (1992). Estimation of levels of gene flow from DNA sequence data. Genetics 132, 583–589.

Janila, P., Pandey, M. K., Shasidhar, Y., Variath, M. T., Sriswathi, M., Khera, P., … Varshney, R. K. (2016). Molecular breeding for introgression of fatty acid desaturase mutant alleles (ahFAD2A and ahFAD2B) enhances oil quality in high and low oil containing peanut genotypes. Plant Science, 242, 203–213. https://doi.org/10.1016/j.plantsci.2015.08.013

Johnsen, P.B., Civille, C.V., Vercellotti, J.R., Sanders, T.H., and Dus, C.A. 1989. Development of a lexicon for the description of peanut flavor. J. Sens. Stud. 3, 9-17.

Jonnala, R. S., Dunford, N. T., & Dashiell, K. E. (2006). Tocopherol, phytosterol and phospholipid compositions of new high oleic peanut cultivars. Journal of Food Composition and Analysis, 19(6–7), 601–605. https://doi.org/10.1016/j.jfca.2006.01.005

Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S. Y., Freimer, N. B., … Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. Nature Genetics, 42(4), 348–354. https://doi.org/10.1038/ng.548

Kaya, C., Tuna, A. L., Dikilitas, M., Ashraf, M., Koskeroglu, S., & Guneri, M. (2009). Supplementary phosphorus can alleviate boron toxicity in tomato. Scientia Horticulturae, 121(3), 284–288. https://doi.org/10.1016/j.scienta.2009.02.011

Kim, M. Y., Lee, S., Van, K., Kim, T. H., Jeong, S. C., Choi, I. Y., *et al.,*(2010). Whole- genome sequencing and intensive analysis of the undomesticated soybean (Glycine soja Sieb. and Zucc.) genome. Proc. Natl. Acad. Sci. U.S.A. 107, 22032–22037. doi: 10.1073/pnas.1009526107

Leclercq, S., Reineccius, G. A., & Milo, C. (2007). Effect of type of oil and addition of δ-tocopherol on model flavor compound stability during storage. Journal of Agricultural and Food Chemistry, 55(22), 9189–9194. https://doi.org/10.1021/jf071686i

Leclercq, S., Reineccius, G. A., & Milo, C. (2007). Model studies on the influence of matrix type and storage environment on the stability of a model aroma mixture during storage. Journal of Agricultural and Food Chemistry, 55(2), 421–425. https://doi.org/10.1021/jf062362h

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lin, T., Zhu, G. T., Zhang, J. H., Xu, X. Y., Du, Y. C., and Huang, S. W. (2014). Genomic analyses provide insights into the history of tomato breeding. Nat. Genet. 46, 1220–1226. doi: 10.1038/ng.3117

Linforth, R. S., & Taylor, A. (2010). Food flavour technology. Chichester, U.K.: Wiley- Blackwell.

Dwivedi, S. L., Bertioli, D. J., Crouch, J. H., Valls, J. F., Upadhyaya, H. D., Fávero, A., *et al.,*(2007). "Peanut," in Genome Mapping and Molecular Breeding in Plants. Oilseeds, Vol. 2, ed C. Kole (Berlin; Heidelberg: Springer-Verlag), 115–151.

Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., *et al.,*(2012). GAPIT: genome association and prediction integrated tool. Bioinformatics 28, 2397–2399. doi: 10.1093/bioinformatics/bts444

Liu, R., Jin, Q., Huang, J., Liu, Y., Wang, X., Mao, W., & Wang, S. (2011). Photodegradation of Aflatoxin B1 in peanut oil. European Food Research and Technology, 232(5), 843–849. https://doi.org/10.1007/s00217-011-1452-6

Liu, Y., Zhao, G., Ren, J., Zhao, M., & Yang, B. (2011). Effect of denaturation during extraction on the conformational and functional properties of peanut protein isolate. Innovative Food Science and Emerging Technologies, 12(3), 375–380. https://doi.org/10.1016/j.ifset.2011.01.012

Lykomitros, D., Fogliano, V., & Capuano, E. (2016). Flavor of roasted peanuts (*Arachis hypogaea*) - Part I: Effect of raw material and processing technology on flavor, color and fatty acid composition of peanuts. Food Research International, 89, 860–869. https://doi.org/10.1016/j.foodres.2016.09.024

Lykomitros, D., Fogliano, V., & Capuano, E. (2016). Flavor of roasted peanuts (*Arachis hypogaea*) - Part II: Correlation of volatile compounds to sensory characteristics. Food Research International, 89, 870–881. https://doi.org/10.1016/j.foodres.2016.08.017

Ma, Q. H. (2010). Functional analysis of a cinnamyl alcohol dehydrogenase involved in lignin biosynthesis in wheat. Journal of Experimental Botany, 61(10), 2735–2744. https://doi.org/10.1093/jxb/erq107

Maguire, L. S., O'Sullivan, S. M., Galvin, K., O'Connor, T. P., & O'Brien, N. M. (2004). Fatty acid profile, tocopherol, squalene and phytosterol content of walnuts, almonds, peanuts, hazelnuts and the macadamia nut. International Journal of Food Sciences and Nutrition, 55(3), 171–178. https://doi.org/10.1080/09637480410001725175

Meilgaard, M., Civille, C.V., and Carr, B.T. 1999. Sensory Evaluation Techniques. 3rd ed. CRC Press, Boca Raton, FL., pp. 231-263.

Montgomery, G. W., Campbell, M. J., Dickson, P., Herbert, S., Siemering, K., Ewen-White, K. R., … Martin, N. G. (2005). Estimation of the rate of SNP genotyping errors from DNA extracted from different tissues. Twin Research and Human Genetics, 8(4), 346–352. https://doi.org/10.1375/1832427054936673

Neta, E. R., Sanders, T. H., & Drake, M. A. (2010). Understanding peanut flavor: A current review. In Y. H. Hui (Ed.), Handbook of fruit and vegetable flavors (pp. 985–1022). NewJersey: John Wiley & Sons Inc.

Ng, E. C., & Dunford, N. T. (2009). Flavour characteristics of peanut cultivars developed forsouthwestern United States. International Journal of Food Science and Technology, 44(3), 603–609.

Nursten H. E. (2005) The Maillard Reaction: Chemistry, Biochemistry, and Implications. books.google.com.

Pak, T. R., Lynch, G. R., & Tsai, P. S. (2002). Estrogen accelerates gonadal recrudescence in photo-regressed male Siberian hamsters. Endocrinology, 143(10), 4131–4134. https://doi.org/10.1210/en.2002-220569

Pandey, M. K., Wang, H., Khera, P., Vishwakarma, M. K., Kale, S. M., Culbreath, A. K., *et al.,*(2017). Genetic dissection of novel QTLs for resistance to leaf spots and tomato spotted wilt virus in peanut (*Arachis hypogaea* L.). Front. Plant Sci. 8:25. doi: 10.3389/fpls.2017.00025

Passone, M. A., Bluma, R., Nesci, A., Resnik, S., & Etcheverry, M. G. (2008). Impact of food grade antioxidants on peanut pods and seeds mycoflora in storage system from Córdoba, Argentina. Journal of Food Safety, 28(4), 550–566. https://doi.org/10.1111/j.1745-4565.2008.00130.x

Pattee, H. E., Giesbrecht, F. G., & Isleib, T. G. (1995). Roasted Peanut Flavor Intensity Variations Among U.S. Genotypes 1 . Peanut Science, 22(2), 158–162. https://doi.org/10.3146/i0095-3679-22-2-16

Pattee, H. E., Isleib, T. G., & Giesbrecht, F. G. (1997). Genotype-by-Environment Interaction in Sweet and Bitter Sensory Attributes of Peanut. Peanut Science, 24(2), 117–123. https://doi.org/10.3146/i0095-3679-24-2-12

Pattee, H. E., Isleib, T. G., & Giesbrecht, F. G. (1998). Variation in Intensity of Sweet and Bitter Sensory Attributes Across Peanut Genotypes 1 . Peanut Science, 25(2), 63–69. https://doi.org/10.3146/i0095-3679-25-2-2

Pattee, H. E., Isleib, T. G., Giesbrecht, F. G., & McFeeters, R. F. (2000). Investigations into genotypic variations of peanut carbohydrates. Journal of Agricultural and Food Chemistry, 48(3), 750–756. https://doi.org/10.1021/jf9910739

Pattee, H. E., Isleib, T. G., Giesbrecht, F. G., & McFeeters, R. F. (2000). Relationships of sweet, bitter, and roasted peanut sensory attributes with carbohydrate components in peanuts. Journal of Agricultural and Food Chemistry, 48(3), 757–763. https://doi.org/10.1021/jf9910741

Pattee, H. E., Isleib, T. G., Gorbet, D. W., Giesbrecht, F. G., & Cui, Z. (2001). Parent Selection in Breeding for Roasted Peanut Flavor Quality 1 . Peanut Science, 28(2), 51–58. https://doi.org/10.3146/i0095-3679-28-2-3

Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nature Genetics, 38(8), 904–909. https://doi.org/10.1038/ng1847

Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2005). NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research, 33(DATABASE ISS.), 501–504. https://doi.org/10.1093/nar/gki025

Qi, H., Zhang, Q., Zhao, T., Chen, R., Zhang, H., Niu, X., & Li, Z. (2005). Antioxidant activity of different sulfate content derivatives of polysaccharide extracted from Ulva pertusa (Chlorophyta) in vitro. International Journal of Biological Macromolecules, 37(4), 195–199. https://doi.org/10.1016/j.ijbiomac.2005.10.008

Sanders, T.H., Vercellotti, J.R., Crippen, K.L., Civille, G.V. (1989). Effect of maturity on roast color and descriptive flavor of peanuts. J. Food Sci. 54(2), 475-477.

Schirack, A. V., Drake, M., Sanders, T. H., & Sandeep, K. P. (2006). Impact of microwave blanching on the flavor of roasted peanuts. Journal of Sensory Studies, 21(4), 428–440. https://doi.org/10.1111/j.1745-459X.2006.00075.x

Seijo, G., Lavia, G. I., Fernández, A., Krapovickas, A., Ducasse, D. A., Bertioli, D. J., *et al.,*(2007). Genomic relationships between the cultivated peanut (*Arachis hypogaea*, Leguminosae) and its close relatives revealed by double GISH. Am. J.Bot. 94, 1963–1971. doi: 10.3732/ajb.94.12.1963

Sontag, T. J., & Parker, R. S. (2007). Influence of major structural features of tocopherols and tocotrienols on their ω-oxidation by tocopherol-ω-hydroxylase. Journal of Lipid Research, 48(5), 1090–1098. https://doi.org/10.1194/jlr.M600514-JLR200

Tseng, Y. C., Tillman, B. L., Peng, Z., and Wang, J. P. (2016). Identification of major QTLs underlying tomato spotted wilt virus resistance in peanut cultivar Florida-EPTM '113'. BMC Genet. 17:128. doi: 10.1186/s12863-016-0435-9

Wang, C. M., Liu, P., Yi, C., Gu, K., Sun, F., Li, L., … Yue, G. H. (2011). A first generation microsatellite- and SNP-based linkage map of jatropha. PLoS ONE, 6(8), 4–11. https://doi.org/10.1371/journal.pone.0023632

Wang, H., Khera, P., Huang, B., Yuan, M., Katam, R., Zhuang, W., … Guo, B. (2016). Analysis of genetic diversity and population structure of peanut cultivars and breeding lines from China, India and the US using simple sequence repeat markers. Journal of Integrative Plant Biology, 58(5), 452–465. https://doi.org/10.1111/jipb.12380

Wang, Z. R., Qiao, S. Y., Lu, W. Q., & Li, D. F. (2005). Effects of enzyme supplementation on performance, nutrient digestibility, gastrointestinal morphology, and volatile fatty acid profiles in the hindgut of broilers fed wheat-based diets. Poultry Science, 84(6), 875–881. https://doi.org/10.1093/ps/84.6.875

Weir, B. S. (1996). Genetic data analysis II: methods for discrete population genetic data. Choice Reviews Online, 34(04), 34-2150-34–2150. https://doi.org/10.5860/choice.34-2150

Xu, Y., & Crouch, J. H. (2008). Marker-assisted selection in plant breeding: From publications to practice. Crop Science, 48(2), 391–407. https://doi.org/10.2135/cropsci2007.04.0191

Zhong, X., Wang, X., Zhou, T., Jin, Y., Tan, S., Jiang, C., … Liu, Z. (2017). Genome-Wide Association Study Reveals Multiple Novel QTL Associated with Low Oxygen Tolerance in Hybrid Catfish. Marine Biotechnology, 19(4), 379–390. https://doi.org/10.1007/s10126-017-9757-5

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nature Genetics, 44(7), 821–824. https://doi.org/10.1038/ng.2310

Zhang, H., Wang, M. L., Schaefer, R., Dang, P., Jiang, T., & Chen, C. (2019). GWAS and Coexpression Network Reveal Ionomic Variation in Cultivated Peanut [Research-article]. Journal of Agricultural and Food Chemistry, 67, 12026–12036. https://doi.org/10.1021/acs.jafc.9b04939

Table 2.1 Distribution of 17,723 SNPs in 20 chromosomes detected in 120 peanut accessions.

Chromosome	Chr length (Mb)	Number of SNPs	Number of genes	Average number of SNPs per Mb	Average number of genes per Mb
A01	112.42	1225	3493	11	31
A02	102.98	709	3129	7	30
A03	143.81	951	4929	7	34
A04	128.8	1059	3833	8	30
A05	115.93	1047	3868	9	33
A06	115.5	880	3712	8	32
A07	64.9	490	2850	8	44
A08	51.9	365	2940	7	57
A09	120.52	415	3694	3	31
A10	117.09	375	3488	3	30
B01	119.44	802	3791	7	32
B02	120.58	742	3580	6	30
B03	117.38	688	5133	6	44
B04	114.59	821	4276	7	37
B05	128.7	727	4433	6	34
B06	123.85	971	4372	8	35
B07	107.94	890	3813	8	35
B08	108.12	1121	3710	10	34
B09	126.9	1658	4409	13	35
B10	115.18	1296	4164	11	36
A genome	1073.85	7516	35936	7	33
B genome	1182.68	8914	41681	8	35
AB genome	2202.01	17232	77617	8	35

Table 2.2 Summary of QTLs associated with Sensory attributes profile, Sugars profile, Fatty Acids profile, and Tocopherols profile.

Traits	QTLs identified	Significant QTLs	-log10 (P-Value)	PVE (%)	MAF
Sensory Attribute Profile					
Sweet	171	13	3.02-4.90	9.93-16.67	0.03-0.12
Bitter	43	1	3.11-4.62	10.24-5.68	0.06-0.42
Sweet Aromatic	63	0	3.01-4.25	9.89-14.37	0.04-0.48
Raw Beany	24	0	3.05-3.99	10.04-13.43	0.05-0.47
Woody Hulls Skins	5	0	3.02-3.30	9.90-10.93	0.05-0.13
Sugars Profile					
Sucrose	175	36	3.01-9.20	9.62-29.93	0.05-0.48
Frucose	103	22	3.01-4.99	9.62-16.56	0.05-0.41
Raffinose	64	15	3.02-5.82	9.65-19.39	0.05-0.48
Inositol	62	0	3.01-4.26	9.63-14.03	0.09-0.45
Total_Sugar	70	14	3.01-6.27	9.63-21.49	0.05-0.46
Fatty Acid Profile					
IV (%)	168	31	3.01-5.63	9.67-17.92	0.08-0.43
Sat (%)	58	4	3.02-4.80	9.64-15.90	0.05-0.45
P/S ratio (%)	164	50	3.01-7.21	9.63-23.85	0.08-0.42
O/L ratio (%)	149	20	3.01-6.64	9.63-22.03	0.06-0.48
LCS (%)	17	0	3.02-3.58	9.67-11.65	0.06-0.44
ROC (%)	13	0	3.17-3.86	10.21-12.64	0.09-0.45
Tocopherols profile					
α-Tocopherol	26	14	3.02-5.70	9.64-18.96	0.05-0.42
β-Tocopherol	90	30	3.04-6.76	9.72-22.41	0.05-0.47
δ-Tocopherol	15	5	3.08-4.68	9.88-15.49	0.05-0.39
γ-Tocopherol	69	13	3.01-5.74	9.63-19.08	0.05-0.41
Total_Tocopherols	30	11	3.06-6.29	9.81-20.92	0.05-0.38

†PVE phenotypic variation explained; MAF minor allele frequency

Table 2.3 The distribution of QTLs significantly associated with four sensory and quality attribute profiles on peanut linkage groups.

Linkage Chr	Flavor			Sugars				
	Sweet	Bitter	Total	Sucrose	Fructose	Raffinose	Total Sugars	Total
A01	0	0	0	1	0	0	0	1
A02	0	0	0	1	0	0	1	2
A03	1	1	2	2	0	0	1	3
A04	3	0	3	6	3	0	4	13
A05	0	0	0	1	1	0	0	2
A06	0	0	0	2	2	0	2	6
A07	0	0	0	0	0	2	0	2
A08	0	0	0	0	0	0	0	0
A09	1	0	1	2	0	1	0	3
A10	0	0	0	0	0	2	0	2
B01	0	0	0	2	0	0	0	2
B02	0	0	0	1	0	0	0	1
B03	0	0	0	2	1	0	1	4
B04	5	0	5	9	5	0	4	18
B05	0	0	0	5	0	0	0	5
B06	3	0	3	0	0	0	0	0
B07	0	0	0	0	0	2	0	2
B08	0	0	0	0	0	0	0	0
B09	0	0	0	1	2	8	1	12
B10	0	0	0	1	0	0	0	1
A_genome	5	1	6	15	6	5	8	34
B_genome	8	0	8	21	8	10	6	45
Total	13	1	14	36	14	15	14	79

Table 2.3 The distribution of QTLs significantly associated with four sensory and quality attribute profiles on peanut linkage groups (continue).

Linkage Chr	Fatty Acid					Tocopherols					
	IV	Sat	PS	OL	Total	Toco alfa	Toco beta	Toco delta	Toco Gamma	Toco Tocal	Total
A01	1	0	3	0	4	0	1	0	0	0	1
A02	3	0	0	0	3	0	2	0	0	0	3
A03	0	0	8	1	9	1	7	0	1	0	9
A04	0	0	1	0	1	1	1	0	1	2	6
A05	4	2	1	3	10	0	3	0	1	0	4
A06	1	0	0	0	1	3	0	0	0	1	5
A07	0	0	0	0	0	0	2	0	0	0	2
A08	1	0	0	0	1	1	0	0	0	1	2
A09	0	0	0	0	0	0	0	0	3	0	3
A10	0	0	0	0	0	0	0	0	0	0	0
B01	1	0	1	0	2	1	3	0	0	0	4
B02	0	0	0	0	0	0	0	2	0	0	2
B03	3	1	22	5	31	0	1	0	0	0	1
B04	0	0	0	0	0	0	0	0	0	0	0
B05	2	0	1	2	5	0	6	0	1	1	8
B06	0	0	0	0	0	7	1	0	1	5	14
B07	0	0	0	0	0	0	1	0	0	0	1
B08	0	0	0	0	0	0	0	0	1	0	1
B09	0	0	0	0	0	0	0	0	4	1	5
B10	15	0	13	9	38	0	2	0	0	0	2
A_genome	10	2	13	4	29	6	16	0	6	4	32
B_genome	21	1	37	16	75	8	14	2	7	7	38
Total	31	3	50	20	104	14	30	2	13	11	70

Table 2.4 The significantly associated SNPs identified for Sensory Attributes profile.

SNP ID	Linkage Chr	Position (bp)	P-Value	- log10 (P value)	PVE (%)	MAF
Sweet						
AX-147248776	14	130,709,725	1.27E-05	4.90	16.67	0.05
AX-176814258	3	12,720,888	2.30E-05	4.64	15.76	0.06
AX-147233387	9	57,329,521	3.09E-05	4.51	15.30	0.49
AX-176817025	4	56,979,926	4.11E-05	4.39	14.87	0.10
AX-176812647	6	103,377,028	4.45E-05	4.35	14.74	0.10
AX-176814709	4	96,322,537	4.49E-05	4.35	14.73	0.10
AX-176807262	14	21,082,467	4.80E-05	4.32	14.63	0.10
AX-176808976	14	29,985,825	4.80E-05	4.32	14.63	0.10
AX-176817147	14	45,705,796	4.80E-05	4.32	14.63	0.10
AX-176810797	14	102,637,105	4.80E-05	4.32	14.63	0.10
AX-176793698	4	84,313,635	4.80E-05	4.32	14.63	0.10
Bitter						
AX-176814258	3	12720888	2.43E-05	4.62	15.68	0.06

†PVE phenotypic variation explained; MAF minor allele frequency

Table 2.5 The significantly associated SNPs identified for Sugars profile.

SNP ID	Linkage Chr	Position (bp)	P-Value	- log10 (P value)	PVE (%)	MAF
Sucorse						
AX-147233387	9	57329521	6.26E-10	9.20	29.93	0.49
AX-176791806	9	57325451	8.30E-10	9.08	29.57	0.06
AX-176814258	3	12720888	7.90E-08	7.10	23.52	0.07
AX-147248776	14	130709725	1.58E-07	6.80	22.56	0.06
AX-176795517	6	107248386	7.83E-07	6.11	20.30	0.05
AX-147249828	15	32289690	2.17E-06	5.66	18.84	0.09
AX-176813185	4	118132903	2.82E-06	5.55	18.45	0.11
AX-176810860	4	103751547	3.23E-06	5.49	18.26	0.10
AX-147222500	5	50695101	3.23E-06	5.49	18.26	0.10
AX-176809744	15	26697331	3.23E-06	5.49	18.26	0.10
AX-176817449	15	31843948	3.23E-06	5.49	18.26	0.10
AX-176802351	4	33883342	3.76E-06	5.42	18.04	0.10
AX-176802397	1	3195706	3.90E-06	5.41	17.98	0.05
AX-176799002	3	15708945	4.07E-06	5.39	17.92	0.10
AX-176817452	15	38756394	5.18E-06	5.29	17.57	0.05
AX-176799580	14	126157122	6.93E-06	5.16	17.14	0.05
AX-176815599	14	3381672	1.07E-05	4.97	16.50	0.06
AX-177639289	20	6469563	1.14E-05	4.94	16.40	0.17
AX-176811232	13	6935401	1.35E-05	4.87	16.15	0.05
AX-176800737	12	3762602	1.69E-05	4.77	15.82	0.06
AX-147259779	19	2420274	1.72E-05	4.77	15.79	0.06
AX-176814709	4	96322537	1.76E-05	4.75	15.76	0.12
AX-176809849	4	97487299	1.99E-05	4.70	15.57	0.12
AX-176812647	6	103377028	2.21E-05	4.66	15.42	0.12
AX-176800704	14	26997129	3.29E-05	4.48	14.82	0.13
AX-176817439	15	28510823	3.56E-05	4.45	14.70	0.05
AX-176793698	4	84313635	3.64E-05	4.44	14.66	0.13
AX-176807262	14	21082467	3.64E-05	4.44	14.66	0.13
AX-176808976	14	29985825	3.64E-05	4.44	14.66	0.13
AX-176817147	14	45705796	3.64E-05	4.44	14.66	0.13
AX-176810797	14	102637105	3.64E-05	4.44	14.66	0.13
AX-176817025	4	56979926	3.70E-05	4.43	14.64	0.13
AX-176810136	2	817431	3.82E-05	4.42	14.59	0.06
AX-147243168	13	3566707	3.82E-05	4.42	14.59	0.06
AX-147237372	11	52532	4.31E-05	4.37	14.41	0.09
AX-147237373	11	52758	4.31E-05	4.37	14.41	0.09
AX-176817084	14	16403547	4.43E-05	4.35	14.37	0.06
Frucose						

AX-176811668	5	91845257	1.02E-05	4.99	16.56	0.10
AX-147226746	6	110928698	1.38E-05	4.86	16.12	0.05
AX-176795330	14	8969353	1.55E-05	4.81	15.95	0.09
AX-176813712	14	9074346	1.55E-05	4.81	15.95	0.09
AX-176817073	14	10046248	1.55E-05	4.81	15.95	0.09
AX-176809746	14	10408809	1.55E-05	4.81	15.95	0.09
AX-176805374	6	109233745	2.60E-05	4.58	15.17	0.10
AX-176812961	4	3782868	2.04E-05	4.69	15.53	0.09
AX-147220028	4	51628148	3.08E-05	4.51	14.92	0.17
AX-176804821	14	54137860	3.12E-05	4.51	14.90	0.08
AX-177641364	19	26148551	3.23E-05	4.49	14.85	0.15
AX-176816801	13	9686512	3.44E-05	4.46	14.75	0.34
AX-176799831	4	45363969	3.77E-05	4.42	14.61	0.17
AX-177644256	19	22490213	4.62E-05	4.34	14.31	0.14
Raffinose						
AX-177637400	19	105708794	1.53E-06	5.82	19.34	0.36
AX-176819101	17	995898	2.04E-06	5.69	18.92	0.41
AX-177639043	19	34765063	7.22E-06	5.14	17.08	0.38
AX-177644524	17	47230439	1.60E-05	4.80	15.90	0.36
AX-176795258	7	62317917	1.60E-05	4.80	15.90	0.36
AX-176820727	19	22420833	1.70E-05	4.77	15.81	0.38
AX-176791962	19	28334973	1.70E-05	4.77	15.81	0.38
AX-177639178	19	25221600	1.94E-05	4.71	15.61	0.38
AX-176797662	9	102772739	1.99E-05	4.70	15.57	0.47
AX-176796351	10	76363955	2.39E-05	4.62	15.30	0.45
AX-177639837	19	18482880	2.41E-05	4.62	15.29	0.38
AX-176806100	7	55457156	2.90E-05	4.54	15.01	0.37
AX-176824170	19	145041727	3.46E-05	4.46	14.74	0.39
AX-176793446	10	77768991	4.58E-05	4.34	14.32	0.45
Total Sugars						
AX-176814258	3	12720888	3.39E-07	6.47	21.49	0.07
AX-176795517	6	107248386	3.65E-06	5.44	18.08	0.05
AX-147248776	14	130709725	6.74E-06	5.17	17.18	0.37
AX-147233387	9	57329521	1.50E-05	4.82	15.99	0.09
AX-176813185	4	118132903	1.55E-05	4.81	15.94	0.11
AX-176799580	14	126157122	1.73E-05	4.76	15.78	0.12
AX-176802351	4	33883342	2.29E-05	4.64	15.36	0.10
AX-176791806	9	57325451	2.38E-05	4.62	15.30	0.06
AX-176812647	6	103377028	2.50E-05	4.60	15.23	0.12
AX-176815599	14	3381672	2.69E-05	4.57	15.12	0.06
AX-147259779	19	2420274	2.88E-05	4.54	15.02	0.06
AX-176814709	4	96322537	2.96E-05	4.53	14.98	0.12
AX-177639289	20	6469563	3.11E-05	4.51	14.90	0.17

AX-176809849	4	97487299	3.86E-05	4.41	14.58	0.12
AX-176810136	2	817431	4.08E-05	4.39	14.49	0.06
AX-147243168	13	3566707	4.08E-05	4.39	14.49	0.06
AX-176817084	14	16403547	4.80E-05	4.32	14.25	0.06

†PVE phenotypic variation explained; MAF minor allele frequency

Table 2.6 The significantly associated SNPs identified for Fatty Acids profile.

SNP ID	Linkage Chr	Position (bp)	P-Value	- log10 (P value)	PVE (%)	MAF
IV (%)						
AX-177640460	20	120772301	2.33E-06	5.63	18.73	0.30
AX-147264395	20	113584807	2.69E-06	5.57	18.53	0.30
AX-177637254	20	119907218	2.69E-06	5.57	18.53	0.30
AX-176804144	20	115730891	2.88E-06	5.54	18.43	0.31
AX-177638677	20	120214136	2.88E-06	5.54	18.43	0.31
AX-177637104	20	121402159	2.88E-06	5.54	18.43	0.31
AX-176819943	15	134187510	4.06E-06	5.39	17.92	0.33
AX-176800147	15	134385876	4.06E-06	5.39	17.92	0.33
AX-176799264	5	93724160	4.06E-06	5.39	17.92	0.33
AX-176804925	20	112467768	4.19E-06	5.38	17.88	0.35
AX-176822799	20	116425340	4.19E-06	5.38	17.88	0.35
AX-177638452	20	117191238	4.52E-06	5.34	17.77	0.16
AX-176792735	5	93859929	6.60E-06	5.18	17.21	0.33
AX-147243963	13	18642606	8.44E-06	5.07	16.85	0.33
AX-176797282	1	6934245	9.42E-06	5.03	16.69	0.14
AX-176822408	13	9113573	1.11E-05	4.96	16.45	0.34
AX-176811155	5	95722033	1.34E-05	4.87	16.17	0.32
AX-176796440	11	10661525	1.55E-05	4.81	15.95	0.32
AX-177637648	20	111329372	1.80E-05	4.74	15.72	0.31
AX-177639314	20	124890936	1.83E-05	4.74	15.70	0.35
AX-176816801	13	9686512	1.89E-05	4.72	15.65	0.34
AX-147264338	20	111357517	2.19E-05	4.66	15.43	0.32
AX-176804911	2	22662900	2.49E-05	4.60	15.23	0.15
AX-176805810	2	22662927	2.49E-05	4.60	15.23	0.15
AX-176791500	20	4946993	2.69E-05	4.57	15.12	0.27
AX-176792724	2	6344117	2.72E-05	4.57	15.10	0.31
AX-147221477	5	1898385	3.87E-05	4.41	14.57	0.31
AX-177639549	20	112215783	3.99E-05	4.40	14.53	0.34
AX-147224423	6	4727732	4.05E-05	4.39	14.50	0.14
AX-177639835	20	126438874	4.59E-05	4.34	14.32	0.30
AX-147231439	8	42766581	4.66E-05	4.33	14.29	0.32
Sat (%)						
AX-176815995	5	83783072	1.60E-05	4.80	15.90	0.21
AX-176822343	5	59674162	2.49E-05	4.60	15.24	0.31

AX-147263936	20	36197991	3.27E-05	4.49	14.83	0.44
P/S ratio (%)						
AX-147243963	13	18642606	6.23E-08	7.21	23.85	0.33
AX-176807302	13	71269320	4.97E-07	6.30	20.95	0.15
AX-176802555	13	85973572	1.50E-06	5.83	19.38	0.32
AX-176800303	13	71966139	1.89E-06	5.72	19.03	0.31
AX-177640460	20	120772301	1.95E-06	5.71	18.99	0.30
AX-176812877	13	50828292	2.65E-06	5.58	18.55	0.32
AX-176794224	13	51144102	2.65E-06	5.58	18.55	0.32
AX-176816920	13	71083056	2.65E-06	5.58	18.55	0.32
AX-176812558	13	73285066	2.65E-06	5.58	18.55	0.32
AX-176798714	1	82680865	2.65E-06	5.58	18.55	0.32
AX-176803950	3	83639278	2.81E-06	5.55	18.46	0.31
AX-176809440	13	70679022	3.04E-06	5.52	18.34	0.31
AX-176798965	3	35912339	3.16E-06	5.50	18.29	0.17
AX-176820627	13	42072598	3.70E-06	5.43	18.06	0.33
AX-176807020	13	73579709	4.00E-06	5.40	17.95	0.31
AX-176804898	13	102270862	4.00E-06	5.40	17.95	0.31
AX-176803007	3	85646533	4.00E-06	5.40	17.95	0.31
AX-176819452	13	49061948	4.44E-06	5.35	17.79	0.16
AX-176817595	13	81114683	4.59E-06	5.34	17.74	0.32
AX-176808699	13	37912119	4.85E-06	5.31	17.66	0.33
AX-176816890	13	38349597	4.85E-06	5.31	17.66	0.33
AX-176802747	4	18450776	4.85E-06	5.31	17.66	0.33
AX-176811636	3	43664456	4.85E-06	5.31	17.66	0.33
AX-176804144	20	115730891	5.12E-06	5.29	17.58	0.31
AX-177638677	20	120214136	5.12E-06	5.29	17.58	0.31
AX-177637104	20	121402159	5.12E-06	5.29	17.58	0.31
AX-176804925	20	112467768	6.06E-06	5.22	17.34	0.35
AX-176822799	20	116425340	6.06E-06	5.22	17.34	0.35
AX-176802309	3	99994285	7.40E-06	5.13	17.04	0.31
AX-177638452	20	117191238	1.06E-05	4.97	16.51	0.16
AX-176802804	3	76469583	1.08E-05	4.96	16.48	0.16
AX-176812632	13	39758074	1.19E-05	4.93	16.35	0.16
AX-176823501	13	14822762	1.22E-05	4.91	16.30	0.35
AX-147264395	20	113584807	1.30E-05	4.89	16.21	0.30
AX-177637254	20	119907218	1.30E-05	4.89	16.21	0.30
AX-176811706	3	32290418	1.55E-05	4.81	15.95	0.33
AX-176796440	11	10661525	1.64E-05	4.78	15.86	0.32

AX-147249143	15	9068762	1.67E-05	4.78	15.84	0.32
AX-176812422	13	80567791	1.82E-05	4.74	15.70	0.32
AX-177639835	20	126438874	1.88E-05	4.73	15.66	0.30
AX-176822464	1	3696555	2.02E-05	4.69	15.55	0.33
AX-176799733	1	7124014	2.32E-05	4.63	15.34	0.32
AX-176812316	13	48865573	2.84E-05	4.55	15.04	0.16
AX-176800673	13	79761955	2.84E-05	4.55	15.04	0.16
AX-176795787	13	121078994	3.08E-05	4.51	14.92	0.34
AX-177637648	20	111329372	3.09E-05	4.51	14.91	0.31
AX-176815442	3	13133357	3.32E-05	4.48	14.81	0.33
AX-177639314	20	124890936	3.43E-05	4.46	14.76	0.35
AX-147264338	20	111357517	3.49E-05	4.46	14.73	0.32
AX-147221477	5	1898385	4.34E-05	4.36	14.40	0.31
O/L ratio (%)						
AX-176819943	15	134187510	2.31E-07	6.64	22.03	0.33
AX-176800147	15	134385876	2.31E-07	6.64	22.03	0.33
AX-176799264	5	93724160	2.31E-07	6.64	22.03	0.33
AX-176822408	13	9113573	5.40E-07	6.27	20.83	0.34
AX-176792735	5	93859929	5.59E-07	6.25	20.78	0.33
AX-147243963	13	18642606	7.37E-07	6.13	20.39	0.33
AX-176816801	13	9686512	5.32E-06	5.27	17.53	0.34
AX-176791500	20	4946993	7.54E-06	5.12	17.01	0.27
AX-176811155	5	95722033	8.44E-06	5.07	16.85	0.32
AX-176796440	11	10661525	1.34E-05	4.87	16.17	0.32
AX-176798965	3	35912339	1.84E-05	4.74	15.69	0.17
AX-177640460	20	120772301	2.34E-05	4.63	15.33	0.30
AX-176820627	13	42072598	2.51E-05	4.60	15.22	0.33
AX-177639314	20	124890936	2.83E-05	4.55	15.05	0.35
AX-176804144	20	115730891	3.13E-05	4.50	14.89	0.31
AX-177638677	20	120214136	3.13E-05	4.50	14.89	0.31
AX-177637104	20	121402159	3.13E-05	4.50	14.89	0.31
AX-176804925	20	112467768	3.83E-05	4.42	14.59	0.35
AX-176822799	20	116425340	3.83E-05	4.42	14.59	0.35
AX-177638452	20	117191238	4.21E-05	4.38	14.45	0.16
AX-176800303	13	71966139	4.49E-05	4.35	14.35	0.31

†PVE phenotypic variation explained; MAF minor allele frequency

Table 2.7 The significantly associated SNPs identified for Tocopherols profile.

SNP ID	Linkage Chr	Position (bp)	P-Value	- log10 (P value)	PVE (%)	MAF
α-Tocopherol						
AX-176819533	16	12671909	1.99E-06	5.70	18.96	0.06
AX-176792655	16	117556196	5.07E-06	5.29	17.60	0.06
AX-176805357	6	94179951	5.07E-06	5.29	17.60	0.06
AX-147252198	16	11622918	5.16E-06	5.29	17.57	0.05
AX-176808373	16	104214231	5.88E-06	5.23	17.38	0.07
AX-176817673	16	105588679	5.88E-06	5.23	17.38	0.07
AX-176794514	16	105915649	5.88E-06	5.23	17.38	0.07
AX-176807852	16	108453558	5.88E-06	5.23	17.38	0.07
AX-176795517	6	107248386	1.24E-05	4.91	16.28	0.05
AX-176810052	8	45093830	1.39E-05	4.86	16.10	0.26
AX-176795785	6	5124210	1.68E-05	4.78	15.83	0.05
AX-176808631	11	22160973	2.66E-05	4.57	15.14	0.06
AX-176814258	3	12720888	3.90E-05	4.41	14.56	0.07
AX-176810057	4	112402040	4.02E-05	4.40	14.51	0.05
β-Tocopherol						
AX-176812053	17	124694409	1.76E-07	6.76	22.41	0.05
AX-176806543	13	7336269	2.16E-07	6.67	22.12	0.08
AX-176794213	15	18362902	3.52E-06	5.45	18.13	0.36
AX-176804110	15	18587237	3.52E-06	5.45	18.13	0.36
AX-176813453	5	17327471	3.52E-06	5.45	18.13	0.36
AX-176807392	5	18917629	3.52E-06	5.45	18.13	0.36
AX-176818540	4	1656177	3.52E-06	5.45	18.13	0.36
AX-176807189	3	112857458	4.42E-06	5.35	17.80	0.10
AX-176800776	15	16143672	4.55E-06	5.34	17.76	0.36
AX-176819938	15	17861943	4.55E-06	5.34	17.76	0.36
AX-176811294	15	18122159	5.74E-06	5.24	17.42	0.36
AX-176796096	5	18596510	7.03E-06	5.15	17.12	0.37
AX-176817367	15	15304406	8.16E-06	5.09	16.90	0.36
AX-147263251	20	6465564	1.79E-05	4.75	15.73	0.18
AX-177639004	20	6469478	1.79E-05	4.75	15.73	0.18
AX-176796595	16	5942908	1.99E-05	4.70	15.57	0.35
AX-176821973	1	10115045	2.48E-05	4.60	15.24	0.16
AX-176804214	3	112856983	3.33E-05	4.48	14.80	0.36
AX-176820936	11	1409056	3.52E-05	4.45	14.72	0.16
AX-176818525	11	58428809	3.63E-05	4.44	14.67	0.48
AX-176795105	7	73246538	3.90E-05	4.41	14.56	0.35
AX-176814383	3	64446273	3.90E-05	4.41	14.56	0.18
AX-176810428	7	73297977	4.19E-05	4.38	14.45	0.35

AX-176810550	3	1930218	4.81E-05	4.32	14.24	0.18
AX-176817109	3	112856948	4.81E-05	4.32	14.24	0.18
AX-176815701	3	112857177	4.81E-05	4.32	14.24	0.35
AX-176808255	3	112857201	4.81E-05	4.32	14.24	0.35
AX-176795366	2	6392300	4.81E-05	4.32	14.24	0.35
AX-176796733	2	6394704	4.81E-05	4.32	14.24	0.18
AX-176791718	11	1411071	5.05E-05	4.30	14.17	0.17
δ-Tocopherol						
AX-147241099	12	22562082	2.11E-05	4.68	15.49	0.06
AX-176791511	12	30172328	2.11E-05	4.68	15.49	0.06
γ-Tocopherol						
AX-176805798	16	95233197	1.83E-06	5.74	19.08	0.06
AX-147257537	18	7765731	2.15E-06	5.67	18.85	0.05
AX-176817459	15	41204780	5.49E-06	5.26	17.48	0.12
AX-176803682	4	37436565	9.88E-06	5.01	16.61	0.11
AX-176791660	19	143735714	1.16E-05	4.94	16.38	0.31
AX-177638506	19	111617928	1.45E-05	4.84	16.05	0.32
AX-147233375	9	54638085	1.45E-05	4.84	16.05	0.32
AX-147234334	9	113662975	1.58E-05	4.80	15.92	0.45
AX-176796415	9	98176011	2.83E-05	4.55	15.04	0.35
AX-176812851	5	79572378	3.34E-05	4.48	14.80	0.33
AX-176814440	3	106948006	3.48E-05	4.46	14.73	0.33
AX-147262354	19	143694732	3.71E-05	4.43	14.64	0.41
AX-176822955	19	114117157	3.84E-05	4.42	14.58	0.32
Total Tocopherols						
AX-176803682	4	37436565	5.08E-07	6.29	20.92	0.11
AX-176792655	16	117556196	1.80E-06	5.74	19.10	0.06
AX-176805357	6	94179951	1.80E-06	5.74	19.10	0.06
AX-176808373	16	104214231	2.58E-06	5.59	18.59	0.07
AX-176817673	16	105588679	2.58E-06	5.59	18.59	0.07
AX-176794514	16	105915649	2.58E-06	5.59	18.59	0.07
AX-176807852	16	108453558	2.58E-06	5.59	18.59	0.07
AX-176815360	4	22606633	1.88E-05	4.73	15.66	0.10
AX-176817459	15	41204780	2.15E-05	4.67	15.46	0.12
AX-176791660	19	143735714	3.13E-05	4.50	14.89	0.31
AX-176810052	8	45093830	4.83E-05	4.32	14.24	0.26

†PVE phenotypic variation explained; MAF minor allele frequency

Table 2.8 Candidate genes within the potential QTL regions significantly associated with Sensory attributes profile.

Linkage Chr	Gene name	SNP ID	SNP position (bp)	Function or Biological process
Sweet				
B4	*ACJ5YK*	AX-147248776	130709725	cinnamyl alcohol dehydrogenase; oxidoreductase activity
B4	*FMT16N*	AX-147248776	130709725	transferring glycosyl group transferase
B4	*7P4FH0*	AX-147248776	130709725	probable polygalacturonase
B4	*VJ6UAM*	AX-176810797	102637105	Aldehyde/histidinol dehydrogenase
B4	*FDLX2Z*	AX-176810797	102637105	Glycosyl transferase
B4	*42PK2U*	AX-176807262	21082467	Fasciclin-like arabinogalactan family protein
A3	*2X1NQK*	AX-176814258	12720888	Glycosyl transferase; acetylglucosaminyltransferase activity
A4	*3PMI9T*	AX-176817025	56979926	sterol 3-beta-glucosyltransferase
A4	*FW69P6*	AX-176793698	84313635	Glucose/ribitol dehydrogenase
A6	*31KTN8*	AX-176812647	103377028	Galactose oxidase; kelch repeat superfamily protein
Bitter				
A3	*2X1NQK*	AX-176814258	12720888	Glycosyl transferase; acetylglucosaminyltransferase activity
A3	*EZ9WFK*	AX-176814258	12720888	strictosidine synthase activity
A3	*G490LK*	AX-176814258	12720888	FASCICLIN-like arabinogalactan protein
Sweet Aromatic				
A3	*FG250G*	AX-176804432	1475928	Nucleotide/sugar transporter family protein
A3	*2X39TD*	AX-176804432	1475928	Glycoside hydrolase
A3	*77T7YG*	AX-176804432	1475928	Galactose oxidase

Table 2.9 Candidate genes within the QTL regions significantly associated with Sugars profile.

Linkage Chr	Gene name	SNP ID	SNP position (bp)	Function or Biological process
Total Sugars				
A6	*G897EA*	AX-176795517	107248386	xylan alpha-glucuronosyltransferase 3-like
A6	*070EBZ*	AX-176795517	107248386	Sugar transporter SWEET; SWEET sugar transporter
B4	*7P4FH0*	AX-147248776	130709725	probable polygalacturonase; carbohydrate process
A4	*L4K1UG*	AX-176813185	118132903	probable polygalacturonase; carbohydrate process
B4	*THTJ31*	AX-176799580	126157122	NAD-dependent malic enzyme; malate process
B4	*2H46NI*	AX-176799580	126157122	Glucose/ribitol dehydrogenase; oxidoreductase activity
B4	*L4J78V*	AX-176799580	126157122	Glucose/ribitol dehydrogenase; oxidoreductase activity
B4	*0R79PA*	AX-176799580	126157122	Glucose/ribitol dehydrogenase; mannose process
A4	*HI5XRQ*	AX-176802351	33883342	polygalacturonase; carbohydrate metabolic process
A4	*S10LH7*	AX-176802351	33883342	polygalacturonase; carbohydrate metabolic process
A6	*I7CGDJ*	AX-176812647	103377028	D-ribulose-5-phosphate-3-epimerase
A6	*PV44JT*	AX-176812647	103377028	glyceraldehyde-3-phosphate dehydrogenase
B4	*EB8TWM*	AX-176815599	3381672	UDP-D-glucuronate 4-epimerase
B4	*B363MR*	AX-176815599	3381672	sucrose phosphate synthase; sucrose metabolic process
B4	*5E5ZG3*	AX-176815599	3381672	Glycoside hydrolase family 17; carbohydrate process
B9	*36XQKP*	AX-147259779	2420274	carbohydrate metabolic process
A2	*HLUA2X*	AX-176810136	817431	Malic oxidoreductase; malate metabolic process
A2	*BG060F*	AX-176810136	817431	polygalacturonase; carbohydrate metabolic process
B4	*R869CP*	AX-176817084	16403547	beta-xylosidase; carbohydrate metabolic process
B4	*QJW53T*	AX-176817084	16403547	polygalacturonase activity; carbohydrate process
B4	*7I8QVG*	AX-176817084	16403547	polygalacturonase activity; carbohydrate process
Sucrose				
A1	*QR17ZP*	AX-176802397	3195706	carbohydrate metabolic process; glycogen process
A2	*HLUA2X*	AX-176810136	817431	Malic oxidoreductase; malate metabolic process
A2	*BG060F*	AX-176810136	817431	polygalacturonase; carbohydrate metabolic process
A4	*L4K1UG*	AX-176813185	118132903	probable polygalacturonase; carbohydrate process
A4	*T6IMA7*	AX-176813185	118132903	sterol 3-beta-glucosyltransferase; carbohydrate process
A4	*HI5XRQ*	AX-176802351	33883342	polygalacturonase; carbohydrate metabolic process
A4	*S10LH7*	AX-176802351	33883342	polygalacturonase; carbohydrate metabolic process
A5	*9E7XBZ*	AX-147222500	50695101	glucose-6-phosphate dehydrogenase
A6	*070EBZ*	AX-176795517	107248386	Sugar transporter SWEET; SWEET sugar transporter
A6	*PV44JT*	AX-176812647	103377028	glyceraldehyde-3-phosphate dehydrogenase
B1	*74FNJK*	AX-147237372	52532	glucose-6-phosphate dehydrogenase
B1	*RSDZ0N*	AX-147237372	52532	Galactose mutarotase-like domain; mannose process

B2	*RUYM47*	AX-176800737	3762602	polygalacturonase; carbohydrate metabolic process
B2	*7E9Q5B*	AX-176800737	3762602	sugar porter (SP) family MFS transporter
B4	*7P4FH0*	AX-147248776	130709725	probable polygalacturonase; carbohydrate process
B4	*THTJ31*	AX-176799580	126157122	Malic oxidoreductase; malate metabolic process
B4	*2H46NI*	AX-176799580	126157122	Glucose/ribitol dehydrogenase; metabolic process
B4	*L4J78V*	AX-176799580	126157122	Glucose/ribitol dehydrogenase; metabolic process
B4	*0R79PA*	AX-176799580	126157122	mannose biosynthetic process
B4	*EB8TWM*	AX-176815599	3381672	UDP-D-glucuronate 4-epimerase
B4	*B363MR*	AX-176815599	3381672	sucrose phosphate synthase; ucrose metabolic process
B4	*YH79ID*	AX-176817147	45705796	Malate dehydrogenase; oxidation-reduction process
B4	*QJW53T*	AX-176817084	16403547	probable polygalacturonase-like
B5	*58BRI5*	AX-176809744	26697331	isocitrate dehydrogenase; isocitrate metabolic process
B5	*IE214W*	AX-176809744	26697331	1-deoxy-D-xylulose 5-phosphate synthase
B5	*IAPX52*	AX-176817439	28510823	UDP-glucosyltransferase; metabolic process
B9	*36XQKP*	AX-147259779	2420274	Concanavalin A-like lectin/glucanase subgroup
B9	*20WGQW*	AX-147259779	2420274	Concanavalin A-like lectin/glucanase subgroup
Fructose				
A4	*96HG27*	AX-176812961	3782868	1-deoxy-D-xylulose 5-phosphate synthase; terpenoid biosynthetic process
A4	*5QB4T4*	AX-176812961	3782868	sugar transporter; transmembrane transporter activity
A5	*6EUV2M*	AX-176811668	91845257	aldose 1-epimerase family protein; isomerase activity
A6	*SB83T1*	AX-147226746	110928698	beta-fructofuranosidase cell wall invertase
A6	*31H9ZG*	AX-147226746	110928698	beta-fructofuranosidase cell wall invertase fructosidase
A6	*LG83KY*	AX-176805374	109233745	galactosyltransferase activity
B4	*0ET7PE*	AX-176817073	10046248	polygalacturonase; carbohydrate metabolic process
B4	*U11QFP*	AX-176817073	10046248	glucose 6-phosphate/phosphate translocator
Raffinose				
B9	*B59APE*	AX-177639178	25221600	probable polygalacturonase; carbohydrate process
B9	*NAIG2Q*	AX-177639178	25221600	Chitinase family protein; cell wall catabolic process
B7	*7NF1VE*	AX-176819101	995898	Nucleotide-sugar transporter; carbohydrate transport
B7	*BW9991*	AX-176819101	995898	monosaccharide transporter; transmembrane transport
B7	*V10QWI*	AX-176819101	995898	isocitrate dehydrogenase; oxidation-reduction process
B7	*XA27J7*	AX-176819101	995898	UDP-Glycosyltransferase superfamily; metabolic process
B7	*I97S71*	AX-176819101	995898	UDP-Glycosyltransferase superfamily; metabolic process
A9	*CHRI9U*	AX-176797662	102772739	Glucose/ribitol dehydrogenase; oxidoreductase activity
Inositol				
B8	*0A1ABD*	AX-177643211	118242210	short-chain dehydrogenase-reductase
A7	*VDE0FL*	AX-147227663	10326430	probable sugar phosphate/phosphate translocator

Table 2.10 Candidate genes within the regions significantly associated with Fatty Acid profile.

Linkage Chr	Gene name	SNP ID	SNP position (bp)	Function or Biological process
IV				
B10	*3ZW7IP*	AX-147264395	113584807	3-hydroxyisobutyryl-CoA hydrolase-like protein
B10	*X66CUM*	AX-176791500	4946993	3-ketoacyl-CoA synthase; fatty acid process
B10	*5IXZ6F*	AX-176791500	4946993	sphingolipid delta desaturase; fatty acid biosynthetic process
B10	*MF24RD*	AX-176791500	4946993	enoyl-CoA hydratase/isomerase family protein
B3	*133ZSG*	AX-176822408	9113573	3-ketoacyl-CoA synthase; fatty acid process
B3	*V2BP3N*	AX-176822408	9113573	3-ketoacyl-CoA synthase; fatty acid process
A6	*ML1PU9*	AX-147224423	4727732	N-acetyltransferase activity
A6	*N2VJYT*	AX-147224423	4727732	Acyl carrier protein-like; fatty acid process
A6	*PM4BFB*	AX-147224423	4727732	fatty-acyl-CoA binding
A5	*FLDD49*	AX-176799264	93724160	3-hydroxyisobutyryl-CoA hydrolase-like protein 5-like isoform X2
A5	*D9NVDI*	AX-147221477	1898385	glycerol-3-phosphate acyltransferase; glycerophospholipid metabolic process
A5	*C7I9DN*	AX-147221477	1898385	4-coumarate: CoA ligase; metabolic process
A2	*K5EEF8*	AX-176792724	6344117	GDSL-like Lipase/Acylhydrolase superfamily protein; hydrolase activity
A2	*W96AKI*	AX-176792724	6344117	GDSL-like Lipase/Acylhydrolase superfamily protein; hydrolase activity
A2	*KFYC81*	AX-176792724	6344117	GDSL-like Lipase/Acylhydrolase superfamily protein; hydrolase activity
A2	*NLZ7CM*	AX-176792724	6344117	GDSL-like Lipase/Acylhydrolase superfamily protein; hydrolase activity
OL				
B10	*X66CUM*	AX-176791500	4946993	3-ketoacyl-CoA synthase 10; fatty acid process
B10	*5IXZ6F*	AX-176791500	4946993	sphingolipid delta desaturase; fatty acid process
B10	*MF24RD*	AX-176791500	4946993	enoyl-CoA hydratase/isomerase family protein;
B3	*133ZSG*	AX-176822408	9113573	3-ketoacyl-CoA synthase ; fatty acid biosynthetic process
B3	*V2BP3N*	AX-176822408	9113573	3-ketoacyl-CoA synthase ; fatty acid process
A5	*FLDD49*	AX-176799264	93724160	3-hydroxyisobutyryl-CoA hydrolase-like protein 5-like isoform X2
PS				
B10	*3ZW7IP*	AX-147264395	113584807	3-hydroxyisobutyryl-CoA hydrolase-like protein
B5	*X6MFLG*	AX-147249143	9068762	Acyl transferase/acyl hydrolase/ lysophospholipase; lipid metabolic process
B3	*TQ3K6M*	AX-176817595	81114683	cholesterol/phospholipid:diacylglycerol acyltransferase; O-acyltransferase activity
B3	*E6IMKI*	AX-176808699	37912119	3-hydroxyisobutyryl-CoA hydrolase-like protein; metabolic process
B3	*WW5SAZ*	AX-176823501	14822762	Acyl-CoA N-acyltransferase; N-acetyltransferase activity
A5	*BI8WMI*	AX-147221477	1898385	glycerol-3-phosphate acyltransferase
A3	*R702JB*	AX-176798965	35912339	3-hydroxyisobutyryl-CoA hydrolase-like protein
A3	*11NJV3*	AX-176798965	35912339	triacylglycerol lipase 2-like isoform X1; lipid process
A3	*DHME93*	AX-176802309	99994285	long-chain acyl-CoA synthetase; metabolic process
A1	*0FVV8A*	AX-176822464	3696555	Glycerophosphoryl diester phosphodiesterase
A1	*MX8U3M*	AX-176799733	7124014	malonyl CoA-acyl carrier transacylase

			SAT	
B3	*133ZSG*	AX-176822408	9113573	fatty acid biosynthetic process
B3	*V2BP3N*	AX-176822408	9113573	fatty acid biosynthetic process
			ROC	
A3	*DFH8S5*	AX-176814232	2814162	fatty acid biosynthetic process
			LCS	
A7	*VZ1TQJ*	AX-176810056	75161338	fatty acid biosynthetic process

Table 2.11 Candidate genes within the QTL regions significantly associated with Tocopherols profile.

Linkage Chr	Gene name	SNP ID	SNP position (bp)	Function or Biological process
Total_Tocopherols				
B6	*308EKY*	AX-176792655	117556196	phosphoenolpyruvate carboxylase; catalytic activity
A8	*QYLY8C*	AX-176810052	45093830	Tyrosine transaminase family; catalytic activity
A8	*DLM7XC*	AX-176810052	45093830	tyrosine aminotransferase; biosynthetic process
α-Tocopherol				
B6	*A8AWVV*	AX-176817673	105588679	pyridoxal kinase; pyridoxal kinase activity
A8	*QYLY8C*	AX-176810052	45093830	Pyridoxal phosphate-dependent transferase; catalytic activity
A8	*DLM7XC*	AX-176810052	45093830	tyrosine aminotransferase; cellular amino acid metabolic process
A4	*ZZ7V32*	AX-176810057	112402040	Cytochrome P450 superfamily protein; oxidation-reduction process
A3	*79DDAD*	AX-176814258	12720888	peptidyl-prolyl cis-trans isomerase CYP40-like isoform X1; protein binding
β-Tocopherol				
B5	*3GHL6G*	AX-176794213	18362902	Cytochrome P450 superfamily protein; oxidation-reduction process
B5	*M1JRW6*	AX-176819938	17861943	Cytochrome P450 superfamily protein; oxidation-reduction process
B3	*CN4LCL*	AX-176806543	7336269	Cellular retinaldehyde binding/alpha-tocopherol transport
B1	*7AY9Y3*	AX-176820936	1409056	Triose-phosphate transporter domain; transporter activity
B1	*FXLQ18*	AX-176820936	1409056	Thiamine pyrophosphate (TPP)-dependent enzyme; pyruvate decarboxylase
A4	*9P7KAZ*	AX-176818540	1656177	Cytochrome P450 superfamily; proteinoxidation-reduction process
A3	*36ZVTJ*	AX-176810550	1930218	Prephenate dehydratase; prephenate dehydratase activity
A1	*IFN97P*	AX-176821973	10115045	Triose hosphate/phosphoenolpyruvate
δ-Tocopherol				
A4	*E313MA*	AX-176803863	6645276	Cytochrome P450 superfamily protein; oxidation-reduction
A2	*989M5Y*	AX-147215190	91666407	pyruvate orthophosphate dikinase; pyruvate metabolic process
γ-Tocopherol				
B9	*53XZTF*	AX-176822955	114117157	Pyruvate kinase family protein;catalytic activity
B8	*U5WKYJ*	AX-147257537	7765731	Cytochrome P450; oxidation-reduction process
B8	*71EM27*	AX-147257537	7765731	Prephenate dehydratase; metabolic process
A9	*WLW17E*	AX-147234334	113662975	Component(E2) of pyruvate dehydrogenase; complex metabolic process
A9	*W3WWNZ*	AX-147234334	113662975	Cytochrome P450 superfamily; protein oxidation-reduction process

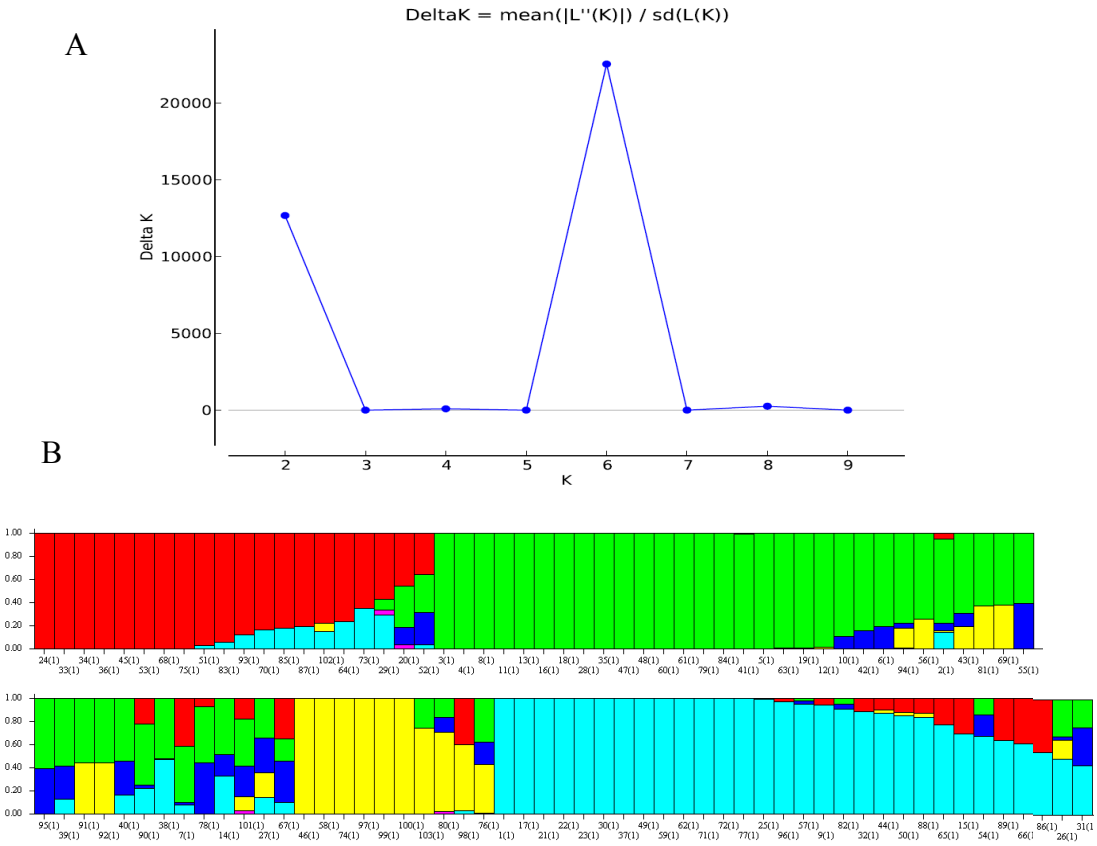


Figure 2.1 Population structure constructed by STRUCTURE analysis. A) Determine best K according to delta K for the SNP dataset. Bottom: Bar plot for K = 6 was created from 120 genotypes from the U.S. peanut Mini-core collection and was ordered by Q values. Each vertical bar represents each genotype and each color represents each cluster.

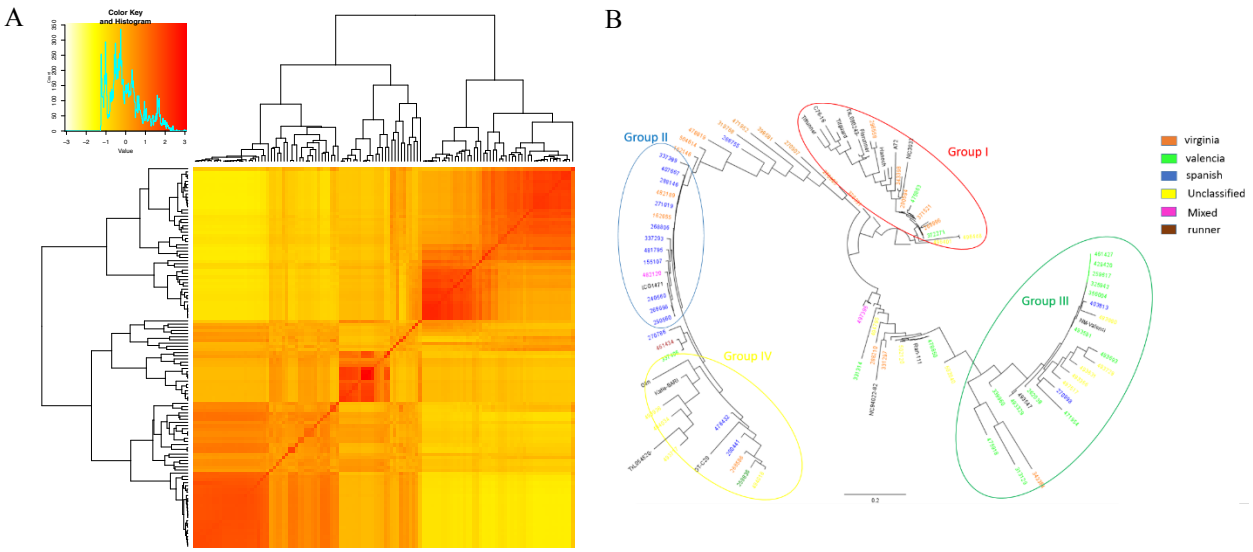


Figure 2.2 Genetic structure of 120 peanut Mini-core collections. A) Kinship Matrix. A heat map of the values in the values in the kinship matrix is created. Kinship Matrix showed that Mini-core collections were grouped into 6 subpopulations. 2) The collections were also clustered by the neighbor-joining tree using whole-genome SNPs. The length of the lines on the tree indicated the simple matching distance. There were four main groups including I: virgina and cultivated peanut; II: spanish; III: valencia and IV: unclassified.

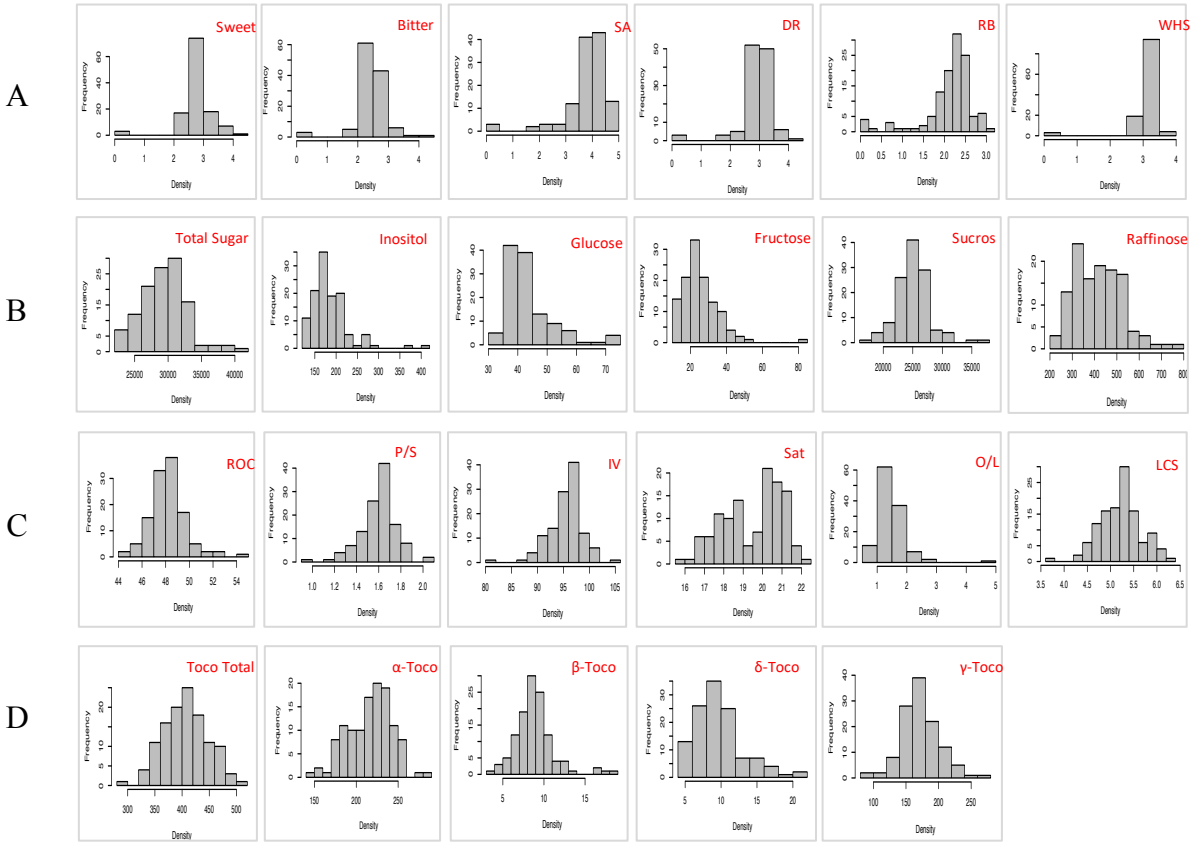


Figure 2.3 Frequency distribution of different sensory and quality attributes in peanut, including A) sensory attributes profile, B) sugars profile, C) fatty acids profile and D) tocopherols profile.

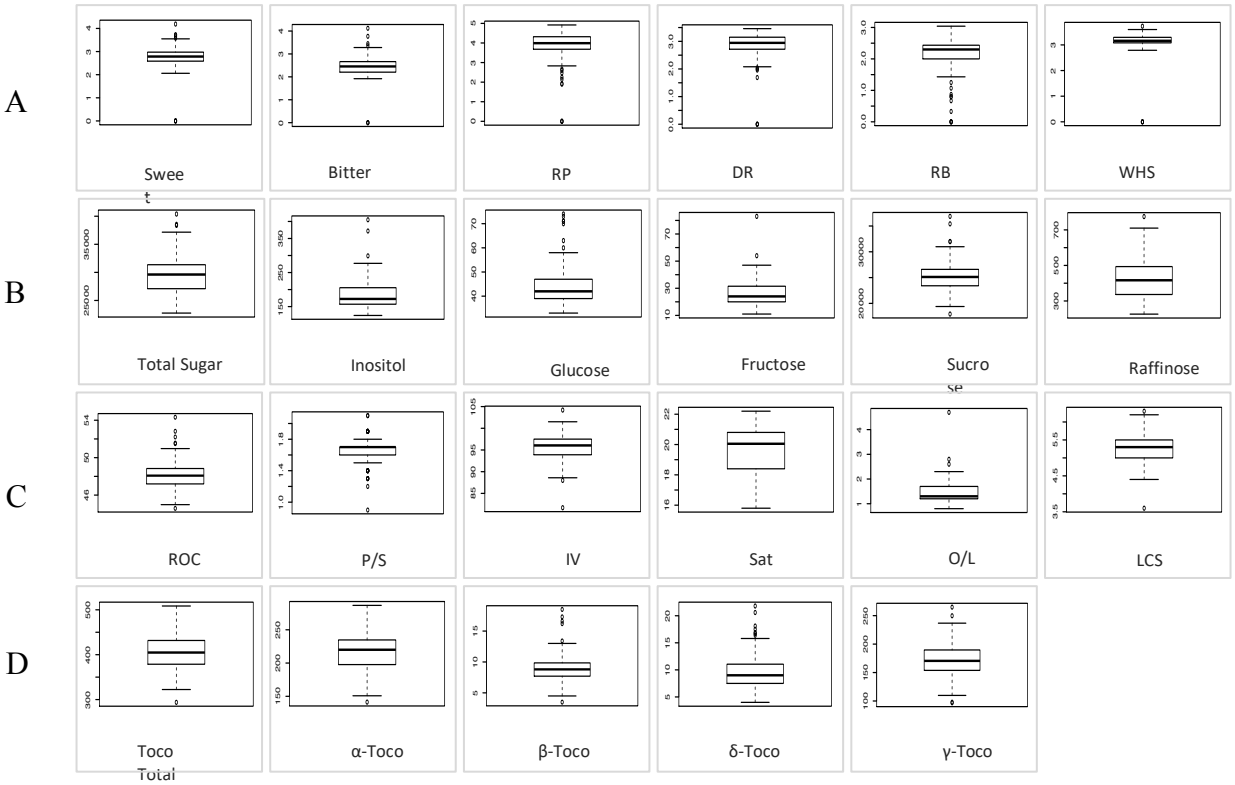


Figure 2.4 Average sensory and quality attribute profile rating scores in peanut Mini-core collections. A) sensory attributes profile, B) sugars profile, C) fatty acids profile and D) tocopherols profile.

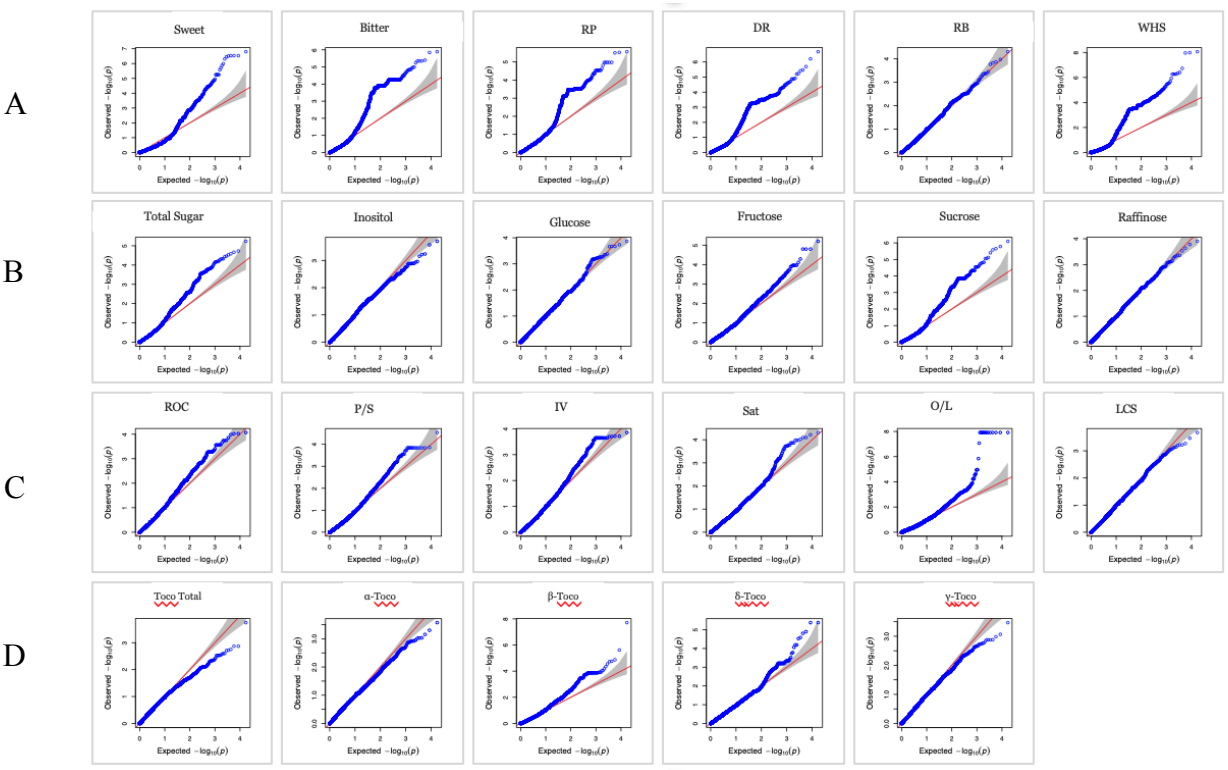


Figure 2.5 Quantile-quantile (QQ) plots of different sensory and quality attributes in peanut, including A) sensory attributes profile, B) sugars profile, C) fatty acids profile and D) tocopherols profile. The Y-axis was the observed negative base 10 logarithm of p-values, and X-axis was the expected observed negative base 10 logarithm of the p-values under the assumption that the p-values follow a uniform [0,1] distribution. The dotted lines showed the 95% confidence interval for the QQ-plot under the null hypothesis of no association between the SNP and the trait.

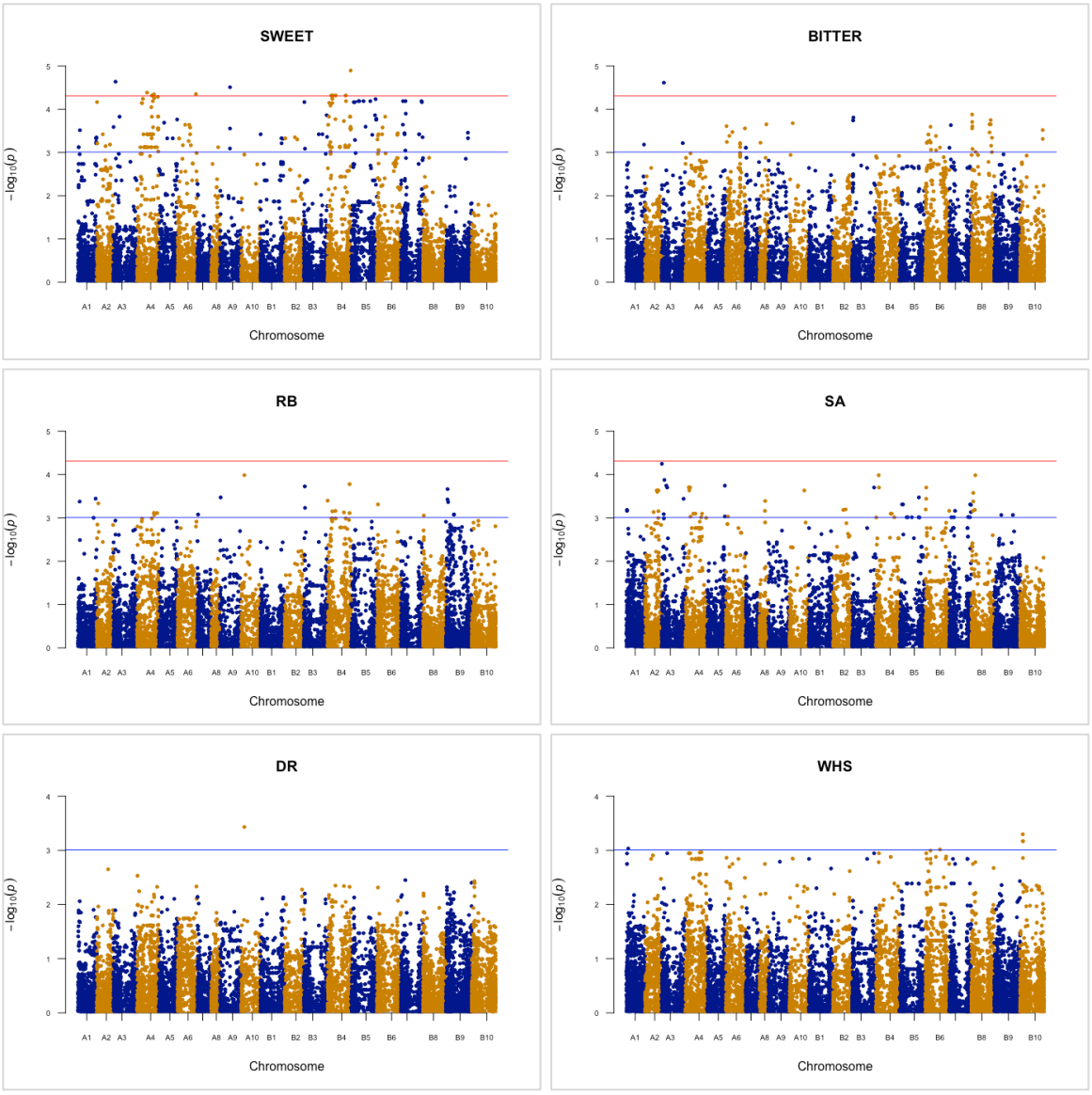


Figure 2.6 Manhattan plots of genome-wide association (GWAS) analysis for peanut sensory attributes profile by EMMAX in SVS analysis. The red solid line represents the genome-wide significant threshold adjusted by Bonferroni-correction: - log10 (P-value) = 4.31. The blue solid line represents the threshold for "suggestive association" adjusted by Bonferroni-correction: - log10 (P-value) = 3.01.

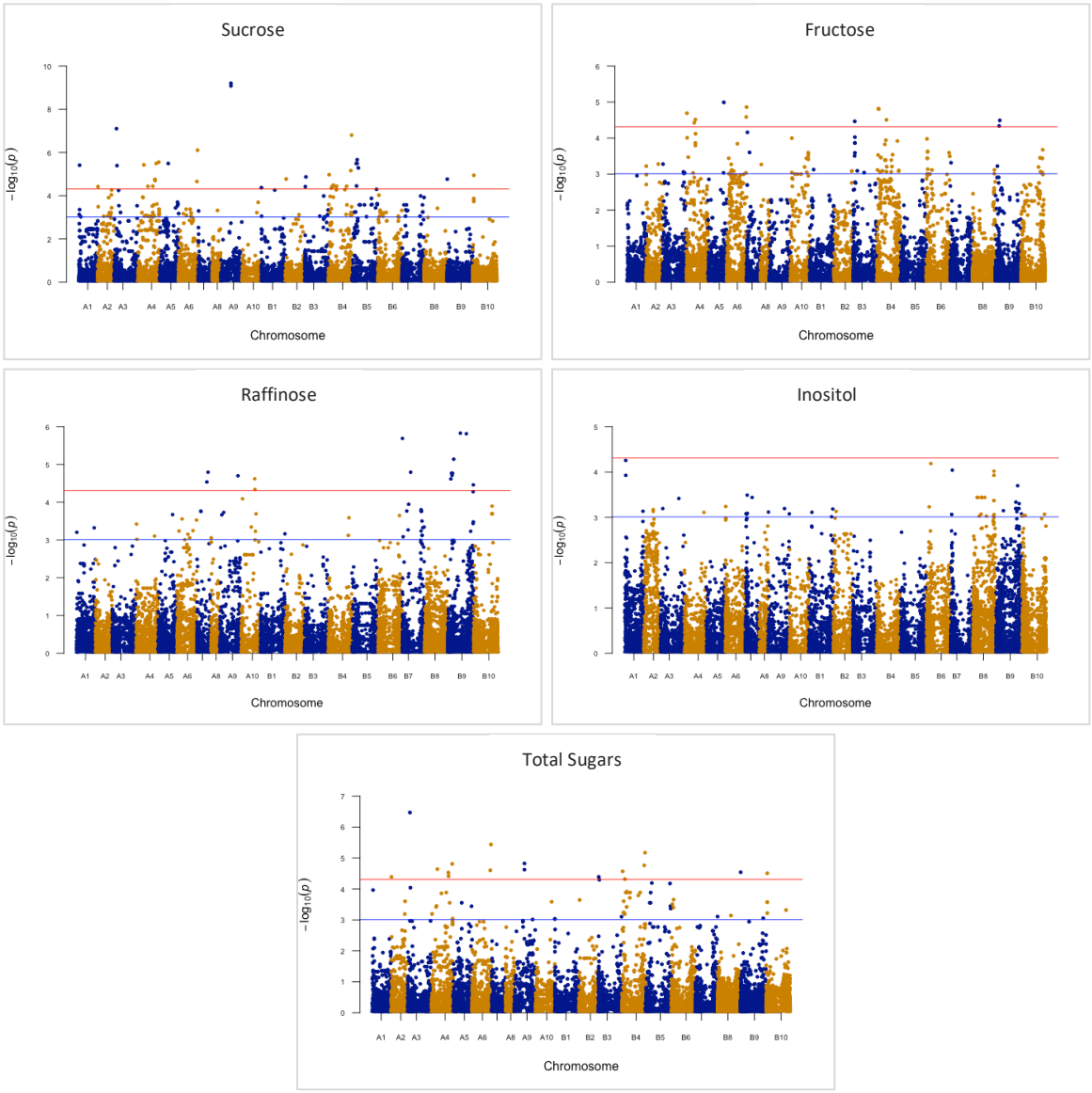


Figure 2.7 Manhattan plots of genome-wide association (GWAS) analysis for peanut sugars profile by EMMAX in SVS analysis. The red solid line represents the genome-wide significant threshold adjusted by Bonferroni-correction: - log10 (P-value) = 4.31. The blue solid line represents the threshold for "suggestive association" adjusted by Bonferroni-correction: - log10 (P-value) = 3.01.

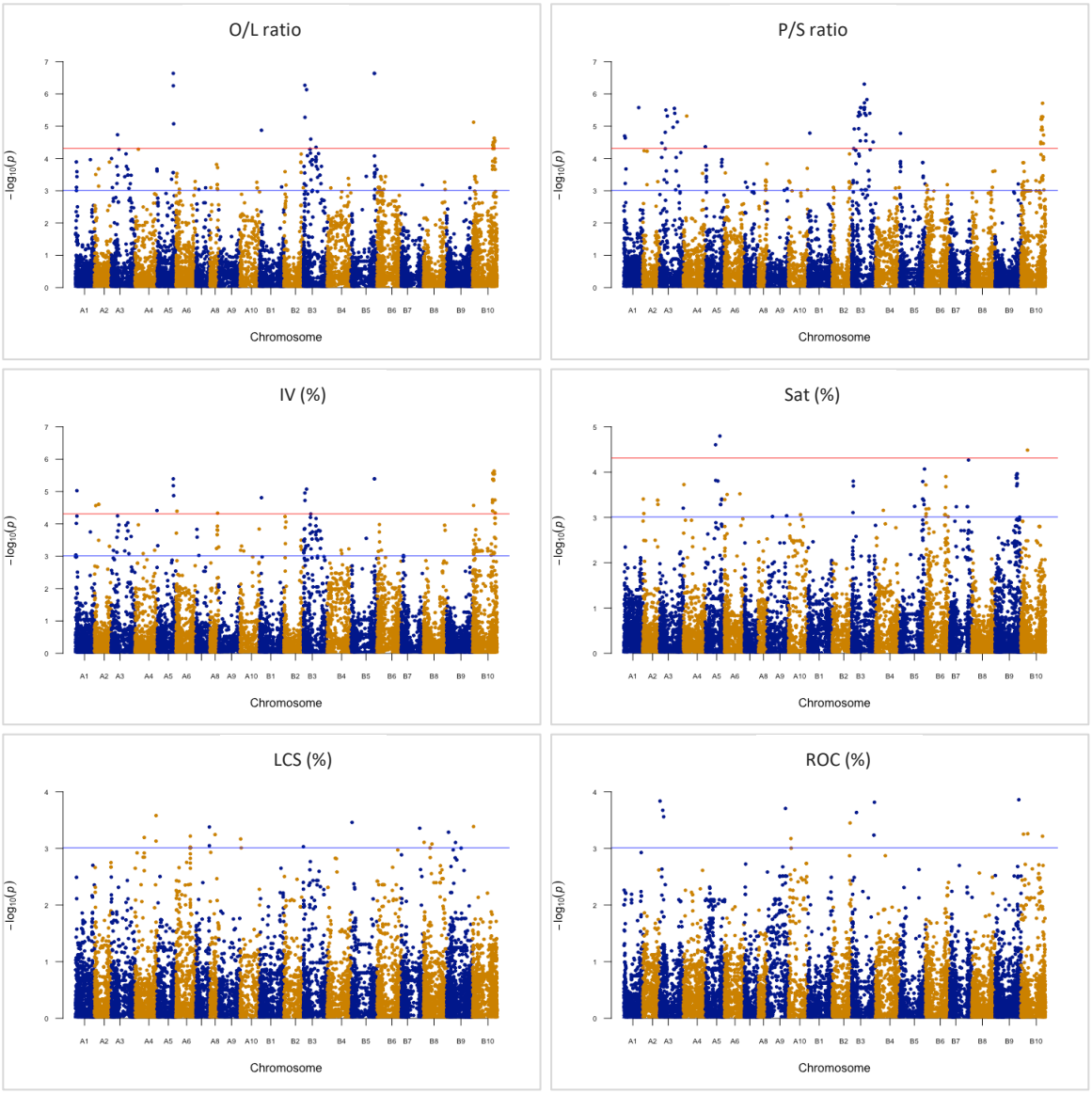


Figure 2.8 Manhattan plots of genome-wide association (GWAS) analysis for peanut fatty acids profile by EMMAX in SVS analysis. The red solid line represents the genome-wide significant threshold adjusted by Bonferroni-correction: - log10 (P-value) = 4.31. The blue solid line represents the threshold for "suggestive association" adjusted by Bonferroni-correction: - log10 (P-value) = 3.01.

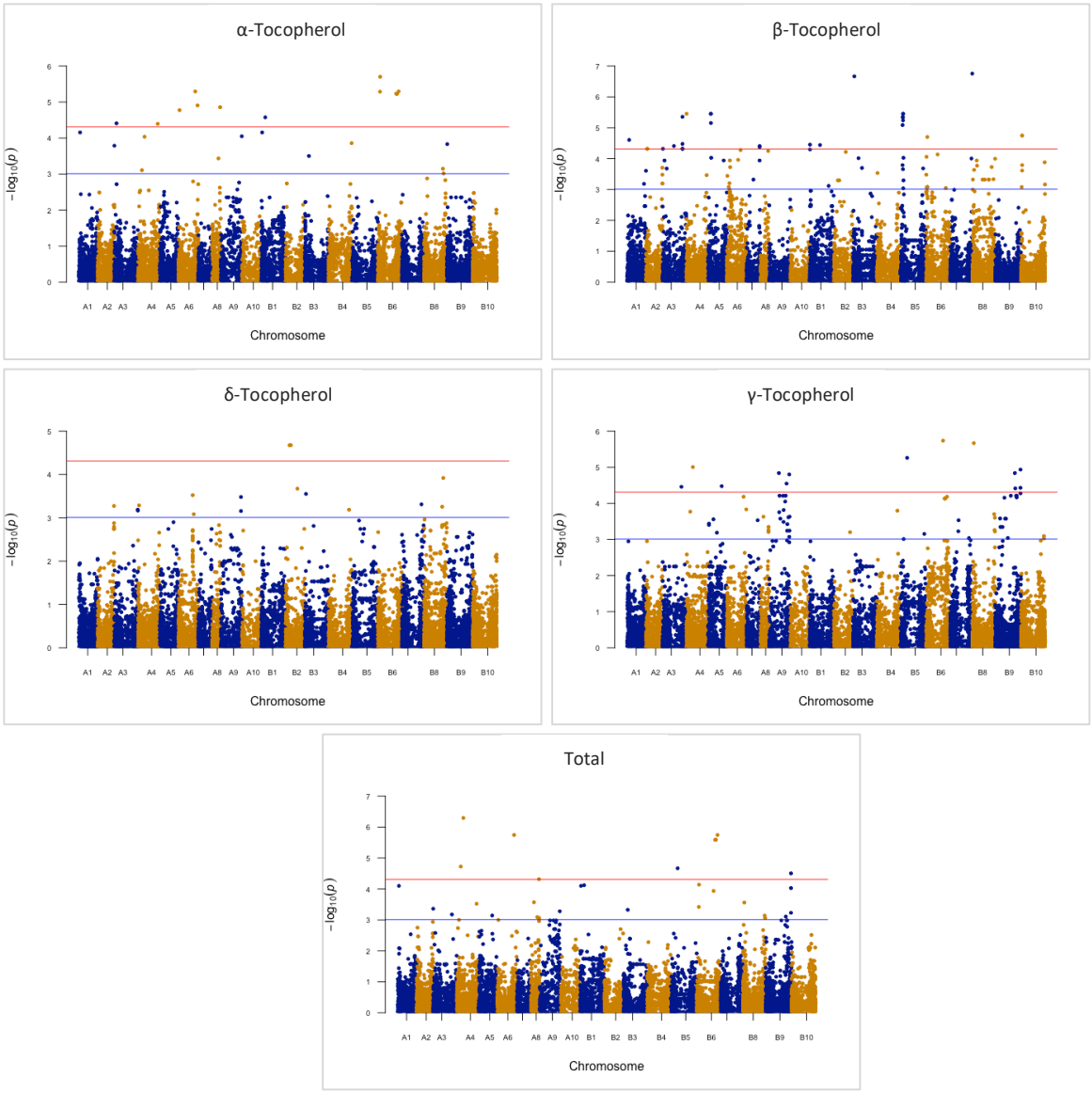


Figure 2.9 Manhattan plots of genome-wide association (GWAS) analysis for peanut tocopherols profile by EMMAX in SVS analysis. The red solid line represents the genome-wide significant threshold adjusted by Bonferroni-correction: - log10 (P-value) = 4.31. The blue solid line represents the threshold for "suggestive association" adjusted by Bonferroni-correction: - log10 (P-value) = 3.01.

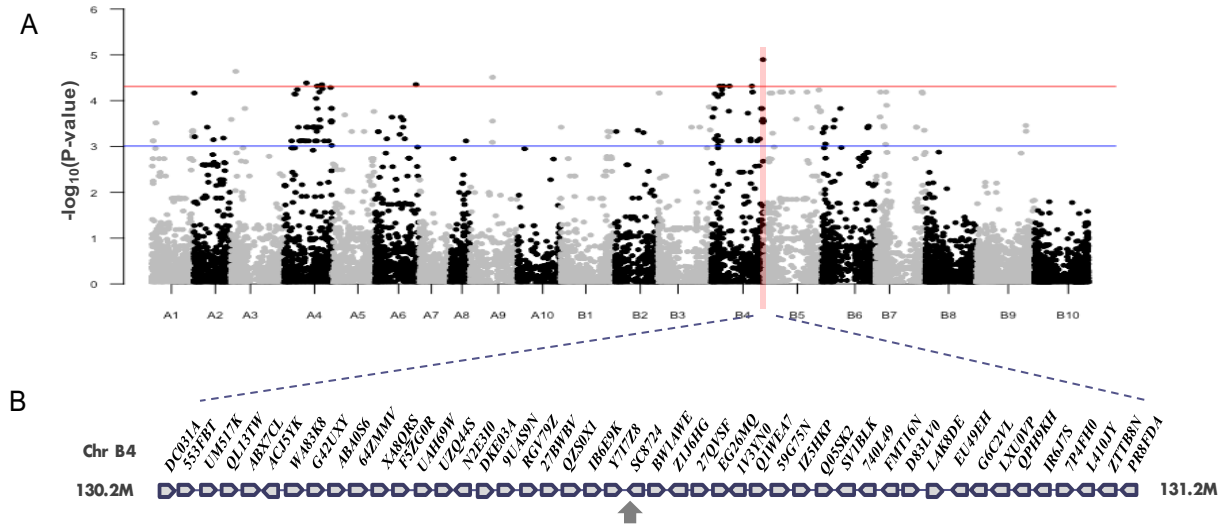


Figure 2.10 Candidate genes around the most significant SNP associated with sensory attributes profile (Sweet trait) on Chromosome B4. A) Zoomed -log10 (P-value) plot for the potential QTLs. The red shade represents ±0.5 Mb genomic region of the most significant SNP. The red solid line represents the genome-wide significant threshold adjusted by Bonferroni-correction: - log10 (P-value) = 4.31. The blue solid line represents the threshold for "suggestive association" adjusted by Bonferroni-correction: - log10 (P-value) = 3.01. B) Candidate genes within 0.5 Mb of both upstream and downstream of the most significant SNP.

Table 2.S1 The distribution of QTLs suggestively associated with sensory attributes profile on peanut linkage groups.

Linkage	Flavor					
Chr	Sweet	Bitter	SA	RB	WHS	Total
A01	9	1	1	2	1	14
A02	5	0	6	1	0	12
A03	3	2	6	0	0	11
A04	47	0	10	2	0	59
A05	4	0	2	0	0	6
A06	10	8	1	0	0	19
A07	0	2	0	1	0	3
A08	1	2	2	0	0	5
A09	3	0	0	1	0	4
A10	0	1	1	1	0	3
B01	4	0	0	0	0	4
B02	3	0	2	0	0	5
B03	6	2	1	2	0	11
B04	37	0	5	6	0	48
B05	15	0	7	0	0	22
B06	10	10	4	1	1	26
B07	12	2	8	0	0	22
B08	0	11	5	1	0	17
B09	2	0	2	6	0	10
B10	0	2	0	0	3	5
A genome	82	16	29	8	1	140
B genome	89	27	34	16	4	170
Total	171	43	63	24	5	310

Table 2.S2 The distribution of QTLs suggestively associated with sugars profile on peanut linkage groups.

Linkage Chr	Sugars					
	Sucrose	Frucose	Raffinose	Inositol	Total Sugar	Total
A01	4	0	2	3	1	10
A02	14	2	0	2	3	21
A03	6	5	0	2	2	15
A04	40	9	3	1	11	64
A05	6	2	1	0	2	11
A06	8	10	6	1	2	27
A07	1	2	4	5	0	12
A08	1	1	1	0	0	3
A09	2	0	3	3	3	11
A10	2	9	6	0	1	18
B01	4	1	1	3	2	11
B02	2	1	0	1	1	5
B03	7	8	0	0	3	18
B04	29	26	2	0	18	75
B05	20	0	0	0	9	29
B06	10	10	1	2	3	26
B07	13	1	16	2	1	33
B08	1	1	0	15	1	18
B09	1	3	12	19	2	37
B10	4	12	6	3	5	30
A genome	84	40	26	17	25	192
B genome	91	63	38	45	45	282
Total	175	103	64	62	70	474

Table 2.S3 The distribution of QTLs suggestively associated with fatty acid profile on peanut linkage groups.

Linkage Chr	Fatty Acid						
	IV	Sat	PS	OL	LCS	ROC	Total
A01	5	0	5	5	0	0	15
A02	5	5	5	5	0	0	20
A03	14	1	19	16	0	3	53
A04	3	1	2	1	3	0	10
A05	5	8	10	8	0	0	31
A06	2	3	1	6	2	0	14
A07	3	0	3	2	2	0	10
A08	7	0	7	7	1	0	22
A09	0	2	2	0	0	1	5
A10	4	1	4	3	1	1	14
B01	1	0	1	2	0	0	4
B02	8	0	8	7	0	1	24
B03	36	3	39	31	1	3	113
B04	3	1	2	3	0	0	9
B05	3	7	10	12	1	0	33
B06	5	13	3	12	0	0	33
B07	1	4	0	1	1	0	7
B08	2	0	3	2	2	0	9
B09	0	8	1	1	2	1	13
B10	61	1	39	25	1	3	130
A genome	48	21	58	53	9	5	194
B genome	120	37	106	96	8	8	375
Total	168	58	164	149	17	13	569

Table 2.S4 The distribution of QTLs suggestively associated with tocopherols profile on peanut linkage groups.

Linkage Chr	Tocopherols					
	Toco alfa	Toco beta	Toco delta	Toco Gamma	Toco Tocal	Total
A01	1	3	0	0	1	5
A02	0	6	1	0	0	7
A03	2	9	2	1	2	16
A04	3	2	1	2	3	11
A05	0	5	0	5	1	11
A06	3	7	2	2	1	15
A07	0	4	0	1	0	5
A08	2	1	0	4	2	9
A09	1	0	2	17	1	21
A10	0	0	0	0	0	0
B01	2	4	0	0	2	8
B02	0	3	3	1	0	7
B03	1	3	1	0	1	6
B04	1	1	1	1	0	4
B05	0	12	0	3	1	16
B06	7	9	0	6	8	30
B07	0	2	0	3	0	5
B08	2	12	2	5	3	24
B09	1	0	0	16	4	21
B10	0	7	0	2	0	9
A genome	12	37	8	32	11	100
B genome	14	53	7	37	19	130
Total	26	90	15	69	30	230

Table 2.S5 One hundred and twenty peanut Mini-core accessions used for GWAS analysis related with Sensory Attribute profiles.

Code	PI no.	Botanical variety	Origin	Code	PI no.	Botanical variety	Origin
G-001	PI152146	*fastigiata*	Uruguay	G-036	PI196622	*hypogaea*	Cote D'Ivoire
G-002	PI158854	*fastigiata*	China	G-037	PI196635	*hypogaea*	Madagascar
G-003	PI200441	*fastigiata*	Japan	G-038	PI240560	*hypogaea*	South Africa
G-004	PI259617	*fastigiata*	Cuba	G-039	PI259658	*hypogaea*	Cuba
G-005	PI259836	*fastigiata*	Malawi	G-040	PI259851	*hypogaea*	Malawi
G-006	PI262038	*fastigiata*	Brazil	G-041	PI268586	*hypogaea*	Zambia
G-007	PI290566	*fastigiata*	India	G-042	PI268696	*hypogaea*	South Africa
G-008	PI290620	*fastigiata*	Argentina	G-043	PI268755	*hypogaea*	Zambia
G-009	PI295730	*fastigiata*	India	G-044	PI268806	*hypogaea*	Zambia
G-010	PI313129	*fastigiata*	Taiwan	G-045	PI268868	*hypogaea*	Sudan
G-011	PI337406	*fastigiata*	Paraguay	G-046	PI268996	*hypogaea*	Zambia
G-012	PI339960	*fastigiata*	Argentina	G-047	PI270786	*hypogaea*	Zambia
G-013	PI343398	*fastigiata*	Israel	G-048	PI270905	*hypogaea*	Zambia
G-014	PI356004	*fastigiata*	Argentina	G-049	PI270907	*hypogaea*	Zambia
G-015	PI429420	*fastigiata*	Zimbabwe	G-051	PI290536	*hypogaea*	India
G-016	PI471954	*fastigiata*	Zimbabwe	G-052	PI290594	*hypogaea*	India
G-017	PI478850	*fastigiata*	Uganda	G-053	PI292950	*hypogaea*	South Africa
G-018	PI482189	*fastigiata*	Zimbabwe	G-054	PI295250	*hypogaea*	Israel
G-019	PI502040	*fastigiata*	Peru	G-055	PI295309	*hypogaea*	Israel
G-020	PI475863	*fastigiata*	Bolivia	G-056	PI296550	*hypogaea*	Israel
G-021	PI475918	*fastigiata*	Bolivia	G-057	PI296558	*hypogaea*	Israel
G-022	PI476025	*fastigiata*	Peru	G-058	PI298854	*hypogaea*	South Africa
G-023	PI493329	*fastigiata*	Argentina	G-059	PI319768	*hypogaea*	Israel
G-024	PI493356	*fastigiata*	Argentina	G-060	PI323268	*hypogaea*	Pakistan
G-025	PI493547	*fastigiata*	Argentina	G-061	PI325943	*hypogaea*	Venezuela
G-026	PI493581	*fastigiata*	Argentina	G-062	PI331297	*hypogaea*	Argentina
G-027	PI493631	*fastigiata*	Argentina	G-063	PI331314	*hypogaea*	Argentina
G-028	PI493693	*fastigiata*	Argentina	G-064	PI337293	*hypogaea*	Brazil
G-029	PI493717	*fastigiata*	Argentina	G-065	PI337399	*hypogaea*	Morocco
G-030	PI493729	*fastigiata*	Argentina	G-066	PI343384	*hypogaea*	Israel
G-031	PI493880	*fastigiata*	Argentina	G-067	PI355268	*hypogaea*	Mexico
G-032	PI493938	*fastigiata*	Argentina	G-068	PI355271	*hypogaea*	Mexico
G-033	PI159786	*hypogaea*	Senegal	G-069	PI370331	*hypogaea*	Israel
G-034	PI162655	*hypogaea*	Uruguay	G-070	PI372271	*hypogaea*	Unknown
G-035	PI162857	*hypogaea*	Sudan	G-071	PI372305	*hypogaea*	Nigeria

TABLE S5 | Continued

Code	PI no.	Botanical variety	Origin	Code	PI no.	Botanical variety	Origin
G-072	PI399581	*hypogaea*	Nigeria	G-097	PI497639	*fastigiata*	Ecuador
G-073	PI442768	*hypogaea*	Zimbabwe	G-098	PI497318	*hypogaea*	Bolivia
G-074	PI461434	*hypogaea*	China	G-099	PI497395	*hypogaea*	Bolivia
G-075	PI471952	*hypogaea*	Zimbabwe	G-100	PI494018	*vulgaris*	Argentina
G-076	PI476636	*hypogaea*	Nigeria	G-101	PI494034	*vulgaris*	Argentina
G-077	PI481795	*hypogaea*	Mozambique	G-102	PI288210	*vulgaris*	India
G-078	PI482120	*hypogaea*	Zimbabwe	G-103	PI371521	*hypogaea*	Israel
G-079	PI494795	*hypogaea*	Zambia	G-104	PI461427	*hypogaea*	China
G-080	PI496401	*hypogaea*	Burkina Faso	G-105	Grif12545	*aequatoriana*	Ecuador
G-081	PI496448	*hypogaea*	Burkina Faso	G-107	Grif14051	*aequatoriana*	Guatemala
G-082	PI504614	*hypogaea*	Colombia	G-108	PI259748	*Hypogaea*	Peru
G-083	PI338338	*peruviana*	Venezuela	G-109	PI390428	*Hypogaea*	Ecuador
G-084	PI502111	*peruviana*	Peru	G-110	PI468250	*Var. hypogaea*	Bolivia
G-085	PI502120	*peruviana*	Peru	G-111	PI497648	*fastigiata*	Ecuador
G-086	PI155107	*vulgaris*	Uruguay	G-112	PI501272	*hypogeae*	Peru
G-087	PI157542	*vulgaris*	China	G-113	PI576613	*hirsuta*	Mexico
G-088	PI270998	*vulgaris*	Zambia	G-114	PI576614	*hirsuta*	Mexico
G-089	PI271019	*vulgaris*	Zambia	G-116	PI576634	*hirsuta*	Mexico
G-090	PI288146	*vulgaris*	India	G-117	PI576636	*hirsuta*	Mexico
G-091	PI290560	*vulgaris*	India	G-118	PI576637	*hirsuta*	Mexico
G-092	PI403813	*vulgaris*	Argentina	G-119	PI648241	*hirsuta*	Ecuador
G-093	PI407667	*vulgaris*	Thailand	G-120	PI648242	*aequatoriana*	Ecuador
G-094	PI478819	*vulgaris*	India	G-121	PI648245	*aequatoriana*	Ecuador
G-095	PI476432	*hypogaea*	Nigeria	G-122	PI648249	*aequatoriana*	Ecuador
G-096	PI497517	*fastigiata*	Brazil	G-123	PI648250	*aequatoriana*	Ecuador

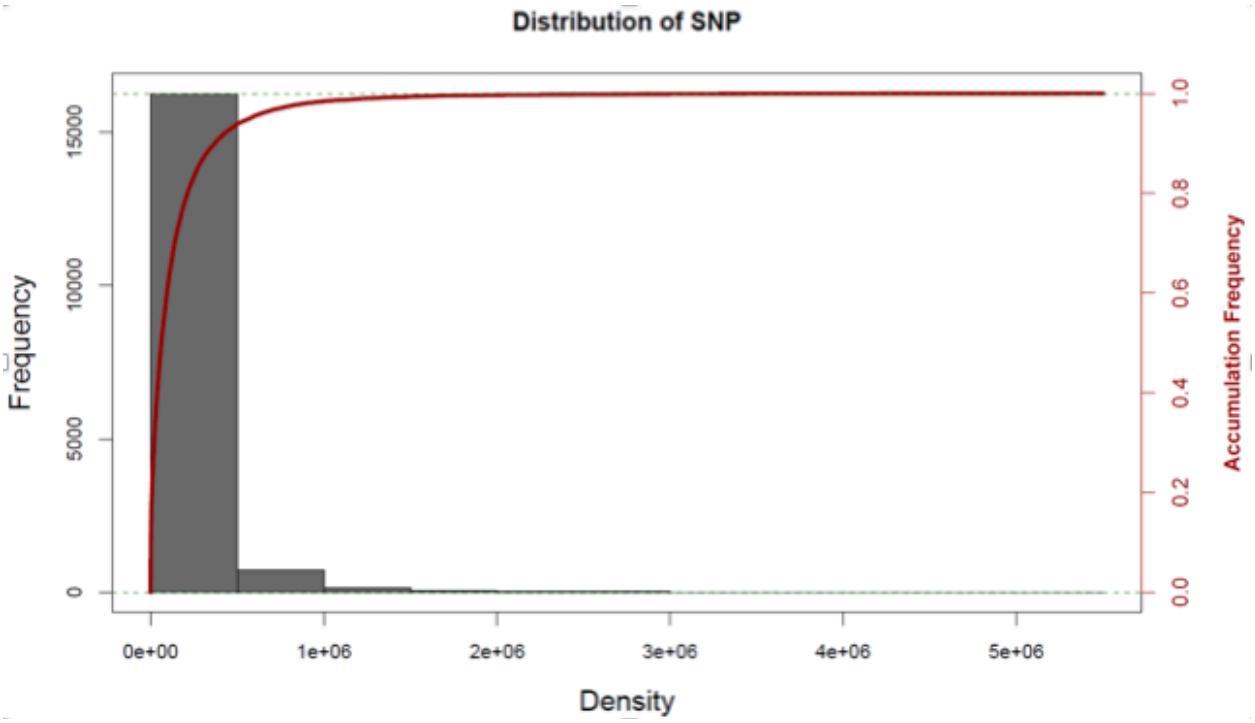


Figure 2.S1 Frequency and accumulative frequency of the selected SNP marker density. Distribution of SNP marker density was showed as a histogram and an accumulative distribution.

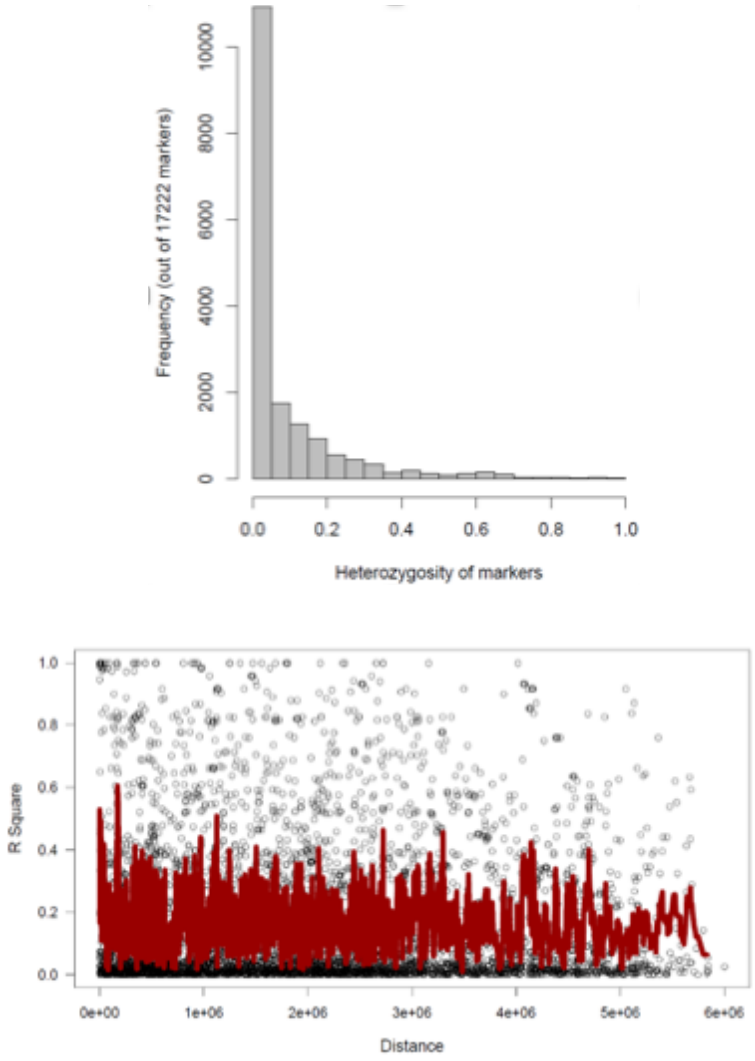


Figure 2.S2 Markers heterozygosity and LD decade. A) The frequency of heterozygous were computed for all of SNP markers. High level of heterozygosis represents low quality. B) Linkage disequilibrium (LD) decade over span. Each dot represented a pair of distance between two markers within the region and their squared correlation coefficient. The red line was the moving average of the 5 adjacent markers.

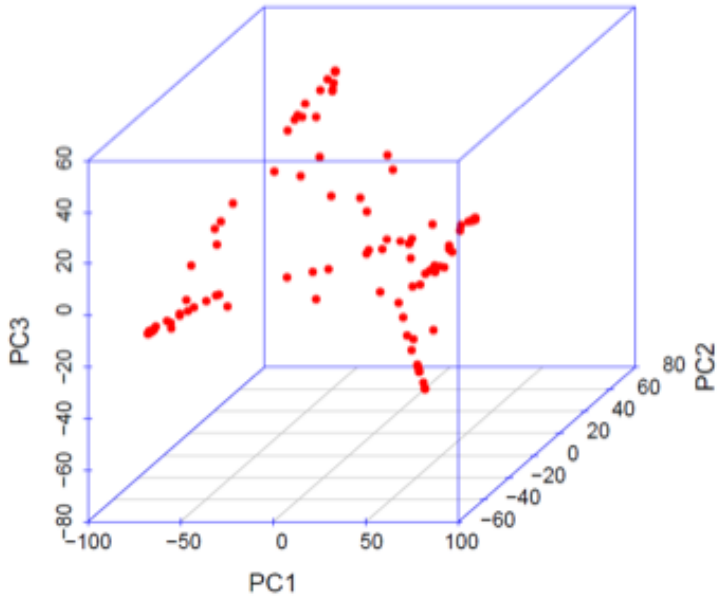


Figure 2.S3 3D plots of principal component of all 120 peanut Mini-core collections.

Chapter Three

SNP marker associated with sensory and quality attributes were confirmed by association mapping using SSRs in peanuts

3.1 Abstract

To determine marker-trait association for seed flavor attributes and seed quality, 104 accessions To determine the marker-trait association for seed flavor and quality attributes, 104 accessions of Mini-core collection was assessed with 131 SSR markers. Seed flavor traits were obtained through a trained sensory assess panel. Researchers often employed the nuclear magnetic resonance (NMR), gas chromatography (GC), and high-performance liquid chromatography (HPLC) analysis to obtain sugars, fatty acid composition, and tocopherols. Genetic diversity and population structure analysis stratified four major subpopulations, which are in accordance with four botanical varieties. According to the association mapping analysis, we know that pPGPseq5D5, GM2745, GM2723, GM1867 are related to sensory attributes, GM1609, Ah32, pPGPseq2C11, XIP297, Ah3 are associated with tocopherols, GM2690, GM2774, GM2791are associated with fatty acids and GM2690, GM1609 are associated with sugars. Prior scholars used Q model, PCA model, Q+K model, and PCA+K model to examine the association between SSR markers and their corresponding traits and always have a consistent high R2 outcome. Of all the SSR markers, GM1609 and GM2690 are commonly used and highly associated markers, which would be used in breeding the flavor-desirable and nutrient-rich cultivars peanuts after further validation is achieved.

3.2 Introduction

Peanut (*Arachis hypogaea* L.), one of the species in the family Fabaceae, belongs to the legume family, which is cultivated through the tropical, subtropical, and warm regions of the world. As

the worldly third-largest oil-seed crop, peanut comprises 50% oil on average, It probably originated in South America, the center of origin, most possibly being Brazil and Peru, in which more than ten wild species were discovered (Acquaah, 2012). Cultivation can produce a spontaneous chromosome duplication hybrid, such as *A. hypogaea* is cross produced by the wild diploid species *A. duranensis* (A-genome) and *A. ipaensis* (B-genome) (Seijo *et al*., 2004). Peanut is genetically diploid, even if it is a tetraploid (AABB) (Stalker *et al*., 1991; Milla et al., 2005).

The construction of genetic linkage maps is a critical step in the quantitative trait locus (QTL) analysis during a marker-assisted selection of plant breeding (Hong et al., 2010). However, lack of sufficient molecular markers hinders current peanut genetic research. Simple sequence repeat (SSR) markers have been the most broadly utilized molecular markers due to their large level of polymorphism, highly abundant, codominant inheritance, practical, readily transferable, and comparatively cost-effective (Vignal et al., 2002). The advantages of SSR over other markers will become more significant when they are used to track desirable traits in large scale breeding programs and as anchor points for map-based gene cloning (He et al., 2003).

Association mapping can easily detect quantitative trait loci (QTL) with relatively large effects on phenotype. It is applied to much broader germplasm than single population trait mapping. In peanut, the first trial at association mapping was published for peanut seed quality traits by using SSR markers and SNP markers in 2011 (Wang et al., 2011). More recently, a genome-wide association study in peanut including 300 genotypes and distinguished 524 significant associations for 36 agronomic traits which could be employed in promoting biotic and abiotic stress resistance, seed sensory and quality, and crop production (Pandey et al., 2014).

However, no study on the SSR marker-traits association in peanut has been reported for sensory and quality attributes in peanut. Therefore, it is essential to carry out association mapping to

identify QTL for sensory and quality attributes in peanut. The objectives of this research were: (1) to examine genetic diversity and population structure in the Mini-core collection by SSR markers; (2) to determine whether the population structure is associated with botanical varieties; (3) to identify whether the employed SSR markers are associated with sensory and quality attributes through association mapping; and (4) to verify the associated SNP markers with sensory and quality attributes obtained from GWAS analysis.

3.3 Materials and methods

3.3.1 Plant material and SSR markers

The experiment included a total of 104 genotypes of *Arachis hypogaea* L. from the U.S. peanut Mini-core collection (Table 2.1), which encompasses four botanical varieties: *fastigiata*, *hypogaea*, *peruviana*, and *vulgaris*.

In view of prior relevant research results, 133 SSR markers in total were applied to genotype the diversity pattern (He et al., 2003, 2005, 2006; Ferguson et al., 2004; Moretzsohn et al., 2005; Budiman et al., 2006; Gimenes, 2007; Cuc et al., 2008; Leal-Bertioli et al., 2009; Nagy et al., 2010; Macedo et al., 2012). Here is the list of the names and references for the 133 polymorphic SSR markers in Table 2.1.

3.3.2 Population structure analysis

The population structure was conducted using program STRUCTURE version 2.2.3 (Pritchard et al., 2000; Falush et al., 2003). STRUCTURE program relies on the model-based clustering and adopts the Bayesian approach to identify cluster complying with Hardy-Weinberg equilibrium and linkage equilibrium. The current study ran the STRUCTURE program for each level of subpopulation (K) value ranging from 2 to 10 with the admixture model for ten times. 20,000 replicates for burn-in, and 20,000 replicates for Markov Chain Monte Carlo (MCMC) procedures

were also involved in each time of programming running. We initially set K values between 2 and 10. Analyses were conducted with correlated allele frequencies (CAF) for the putative populations. The final population subgroups were determined based on the second-order rate of change of the likelihood function (ΔK) (Evanno et al., 2005), which was calculated by using an application on the STRUCTURE Harvester website (Earl and vonHoldt, 2011).

The genetic distances among the subgroups were calculated as Nei's minimum distance and pairwise Fst by PowerMarker. As the STRUCTURE program suggested, we partitioned the genetic variation within and among the populations. PCA was conducted to construct a plot of the most significant axes for validating the genetic structure. The unweighted pair group method with arithmetic average (UPGMA) tree was constructed by MEGA version 6.0 (Tamura et al., 2013) with 100 replications of bootstrapping.

3.3.3 Association mapping analysis

TASSEL version 5.2 was used to conduct the marker-trait association (Bradbury et al., 2007). Four different models with subpopulation membership rate or principle component analysis as fixed covariates and kinship as a random effect were applied in TASSEL (Yu et al., 2006). Q and PCA models adopted the general linear model (GLM): $Y = M\alpha + Qw + e$ (Q model) or $Y = M\alpha + Pw + e$ (PCA model), where Y is phenotypic value, Mα is marker effects, Qw or Pw represents population structure or PCA dimension, and e is the error term. Q+K and PCA+K models adopted mixed linear model (MLM): $Y = M\alpha + Qw + Ku + e$ (Q+K model) or $Y = M\alpha + Pw + Ku + e$ (PCA+K model), where Ku represents familial relatedness.

3.3.4 Alignment analysis between targeted SSRs and SNPs

Genomic survey sequences of targeted SSRs were directly retrieved from Peanut Marker Database at PeanutBase (http://marker.kazusa.or.jp/Peanut/marker/list/Genome-SSR). FASTA files were

obtained based on the genomic survey sequences from NCBI and BLAST, and then they were used to locate the chromosome location of targeted SSR based on the relationship between query and target for the particular BLAST hit at https://peanutbase.org/blast/nucleotide/nucleotide. The targeted SSRs were selected within 3Mb from significant SNPs associated with corresponding traits.

3.4 Results

3.4.1 Population structure analysis

Given ΔK value from STRUCTURE analysis, we chose K = 4 as an optimal grouping (Figure 3.1). Out of 10 runs for K = 4, the run with the highest likelihood value was chosen to designate the subsequent membership coefficients (Q) to every single genotype. A graphical bar plot was created based on Q value (Figure. 3.1). The four subpopulations were assigned with G1, G2, G3, and G4, containing 25, 18, 25, and 36 genotypes, correspondingly (Table 3.2). The fixation index (Fst) with G1 was the highest and followed by G4, G3, and G2, with a value of 0.27, 0.26, 0.22, and 0.04, respectively. Moreover, the genetic distance among these four subpopulations measured by Nei's minimum distance and Fst were consistent with the fixation index result, with the largest genetic distance between G1 and G4 (0.24 and 0.43), and the narrowest genetic distance between G2 and G4 (0.16 and 0.34) (Table 3.3). While the result of principal component analysis (Fig. 3.2) showed that G1 and G3 were well separated from G4, G2 was not entirely separated by principal component 1 (PC1). Also, G1 and G3 were completely separated, but G2 and G4 were not well separated by principal component 2 (PC2). The results from STRUCTURE analysis and principal analysis were consistent.

The UPGMA (unweighted pair group method with arithmetic average) tree analysis clustered the 104 genotypes into four branches, including B1, B2, B3, and B4 (Fig. 3.3) based on the chord distance of SSR markers . Firstly, B1 contained twenty-four genotypes with twenty-three genotypes from subpopulation G1 and one genotype from subpopulation G2. Second, B2 contained two genotypes that were both from subpopulation G2. Third, B3 contained thirty-one genotypes with twenty-three genotypes from subpopulation G3, two genotypes from subpopulations G1, five genotypes from subpopulations G2, and one genotype from subpopulation G4. Finally, B4 contained forty-seven genotypes with thirty-five genotypes from subpopulation G4, one genotype from subpopulation G1, ten genotypes from subpopulation G2, and two genotypes from subpopulation G3. The results from UPGMA tree plot were basically in accord with the conclusions from STRUCTURE plot and PCA analysis except for only several exceptions.

Besides, the 104 genotypes can be classified into four botanical varieties according to the morphological data: *fastigiata*, *hypogaea*, *peruviana*, and *vulgaris*. On the contrary, the four subpopulations were designated based on the SSR allelic variance. From the distribution frequency, 56% of the var. *fastigiata* were classified into subpopulation G1, 67% of the var. *peruviana* were classified into subpopulation G2, 75% of the var. *vulgaris* were classified into subpopulation G3, and 58% of the var. *hypogaea* were classified into subpopulation G4 (Fig.3.4). Despite some discrepancies, the population structure is obviously associated with the botanical variety.

3.4.2 Marker-trait association analysis

Four mixed models were applied to examine the association between the SSR markers and various traits for sensory and quality attributes in peanuts. After the association analysis, SSR markers

were selected based on the p-value and R2 for each model. We selected markers with P-value smaller than 0.05 as markers significantly associated with sensory and quality attributes.

For the flavors profile, four significantly associated SSR markers (pPGPseq5D5, GM2745, GM2723, GM1867) were identified based on all four models, PCA model, Q model, PCA+K model, and Q+K model. Similarly, three SSR markers (GM2690, GM2774, GM2791) were found significantly associated with a fatty acid profile by all four models, while six SSR markers (GM1609, Ah32, pPGPseq2C11, XIP297, Ah3, g10) with tocopherols profile, and two SSR markers (GM2690, GM1609) with sugars profile. While GM2690 is a common marker for the traits of fructose and glucose, indicating certain overlapping of the metabolic process for tocopherols and sugars. The proportion of total variance explained by each significant marker (based on R^2) ranged from 9.40% to 16.03%, with an average R^2 of 12.19%.

3.4.3 SSR markers confirmed by GWAS analysis

A comparison analysis was conducted between significantly SSRs by association mapping and SNPs associated with corresponding sensory and quality attributes by GWAS analysis within 3Mb. We found out that the 37 significantly associated SSRs identified from association mapping were aligned with the QTL SNP markers by GWAS analysis. The 37 significantly associated SSRs included three SSRs associated with flavors profile, 10 SSRs associated with sugars profile, 11 SSRs associated with fatty acids profile, and 13 SSRs associated with the tocopherols profile. Of this list, there were four SNPs (AX-147243168-GM1609, AX-176793698- pPGSseq19A5, AX-176813712- GM1867, AX-147226746- GM2779) significantly associated with sugars profile, one SNP (AX-176804911- pPGPseq1B9) significantly associated with fatty acids profile, and four SNPs (AX-176808373- pPGSseq19A5, AX-147257537- Ah20, AX-176791660- GM1856, AX-

176791511- AC2C05, AX-176805357- TC31H06, AX-176791660- GM1856) significantly associated with tocopherols profile. There were no significant SNPs associated with the flavors profile confirmed by association mapping using SSRs (Table 3.9).

3.5 Discussion

3.5.1 Population structure and genetic diversity analysis

The structure analysis could provide stratification and evaluation of the number of subpopulations, the level of admixture between and among subpopulations, and the genetic relatedness within different varieties. The higher the ΔK value, the higher the similarity within each subpopulation. So, based on the Magnitude of ΔK from STRUCTURE analysis, chose ΔK = 4 was selected as an optimal grouping strategy. In this study, thus, the structure analysis using STRUCTURE stratify the genetic materials into four subpopulations (named as G1, G2, G3, and G4). The chord distance is a measurement of genetic differences, which could be calculated by Powermaker. The lower the value, the closer the distance between genotypes, genetically. Hence, the figure showed that G1 and G3 were well separated from G4 principal component 1, while G2 was not completely separated by principal component 1 (PC1). Also, G1 and G3 were completely separated by principal component 2, but G2 and G4 were not well separated by principal component 2 (PC2). So, we could conclude that the genetic distance between G1 and G4 is farthest, while G2 and G4 are the closest. The genetic distances could also be calculated as Nei's minimum distance and pairwise Fst by PowerMarker, and obtained that the value between G1 and G4 is highest while G2 and G4 is the smallest by both algorithms. The results from STRUCTURE analysis and principal analysis were consistent. Based on the pairwise Fst between and among the subpopulations, the highest diversity was observed between G1 and G4, and the lowest was observed between G2 and

G4. Hence, it could be indicated that G1 and G4 subpopulations have diverged to a greater range as compared to G2 and G4 subpopulations (Table 3.3).

The genetic population structure of the four subpopulations was further validated using the AMOVA analysis, in which 41.56% of the variation could be demonstrated by the stratification into the four subpopulations. This result designated that the separate subpopulations could carry single alleles, and genotypes and the result could be distinguished among taxa, particularly for the more distant taxa, and Cidade et al. also draw the same conclusion (2013). The AMOVA analysis for the botanical variety of peanut germplasm illustrated that 21.21% of the total genetic variance could be ascribed to the discrepancies among botanical species as compared to 16.87% from the earlier results, which divided botanical variety into three subgroups: *hypogaea*, *fastigiata*, and mixed subgroup, combining the other two botanical varieties together (Belamkar et al., 2011). These results support a conclusion from a previous study that botanical variety is a feeble indicator for genetic diversity analysis(Mace et al., 2006).

The membership between genotypes by principal component analysis identified with the relationship by structure analysis. Based on the PCA result, we could conclude that G1, G3, and G4 was almost fully classified by PC1 and PC2, while G2 was scattered along with other subpopulations (Fig. 3.2). The bar plot of structure analysis by STRUCTURE revealed that more than half of the genotypes in G2 were dispersed a large proportion membership compared to other subpopulations (Fig. 3.4). In UPGMA tree plot, the 104 genotypes from U.S. Mini-core collection was gathered into four sections (B1-B4) (Fig. 3.3). Generally speaking, the subpopulations classified by STRUCTURE analysis were consistent with the genetic cluster on the phylogenic analysis, with 96% of B1 coming from G1, all of B2 from G2, 74% of B3 from G3, and 74% of B4 from G4. Furthermore, the subpopulations identified by STRUCTURE analysis were clearly

correlated to the botanical variety (G1~ *fastigiata*, G2~*peruviana*, G3~ *vulgaris*, G4~ *hypogaea*). Nevertheless, only three varieties from *peruviana* were applied in this analysis, including PI502120, PI502111 and, PI338338 in which the first two lines were grouped in subpopulation G2 made up more than 95% proportion of G2. These two lines were also classified into B2 based on UPGMA tree plot. PI338338, collected from Venezuela, was classified in subpopulation G3 with 51% proportion of G3 and 40% proportion of G2, and it was classified in B3 based on UPGMA tree plot. These three genotypes from *peruviana* were separated into two groups using both STRUCTURE and UPGMA analysis, might be caused by a different origin. In UPGMA tree plot, B2 was firstly segregated from the other three branches. This result might be caused by the absence of genotypes from *peruviana*. Besides, B4 (belongs to G4 and *hypogaea*) was segregated from B1 (belongs to G1 and *fastigiata*) and B3 (belongs to G3 and *vulgaris*). These stratification corresponded to the classification for botanical variety, because botanical variety var. *hypogaea* are in the subspecies of *hypogaea*, while botanical variety *fastigiata* and *vulgaris* are in the subspecies of *fastigiata*. Hence, there is a prominent association among botanical variety, population structure, and stratification distinguished by distance-based methods.

3.6 References

Acquaah, G. 2012. Principles of Plant Genetics and Breeding: Second Edition.

Belamkar, V., M.G. Selvaraj, J.L. Ayers, P.R. Payton, N. Puppala, and M.D. Burow. 2011. A first insight into population structure and linkage disequilibrium in the US peanut minicore collection. Genetica 139(4): 411–429.

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(3): 314–31.

Cuc, L.M., E.S. Mace, J.H. Crouch, V.D. Quang, T.D. Long, and R.K. Varshney. 2008. Isolation and characterization of novel microsatellite markers and their application for diversity assessment in cultivated groundnut (*Arachis hypogaea*). BMC Plant Biol. 8(1): 55.

Dang, P.M., and C.Y. Chen. 2013. Modified method for combined DNA and RNA isolation from peanut and other oil seeds. Mol. Biol. Rep. 40(2): 1563–1568.

Earl, D.A., and B.M. vonHoldt. 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4(2): 359–361.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. 14(8): 2611–2620.

Falush, D., M. Stephens, and J.K. Pritchard. 2003. Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. Genetics 164(4): 1567–1587.

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(6): 1064–70.

Gimenes, M.A., Hoshino, A.A., Barbosa, A.V., Palmieri, D.A. and Lopes, C.R., 2007. Characterization and transferability of microsatellite markers of the cultivated peanut (*Arachis hypogaea*). BMC Plant Biology, 7(1), p.1.

He, G., F.E. Woullard, I. Marong, and B.Z. Guo. 2006. Transferability of Soybean SSR Markers in Peanut (*Arachis hypogaea* L.). Peanut Sci. 33(1): 22–28.

Herbert, J. J., & Toews, M. D. (2011). Seasonal Abundance and Population Structure of Brown Stink Bug (Hemiptera: Pentatomidae) in Farmscapes Containing Corn, Cotton, Peanut, and

Soybean. Annals of the Entomological Society of America, 104(5), 909–918. https://doi.org/10.1603/an11060

Hong, Y., Chen, X., Liang, X., Liu, H., Zhou, G., Li, S., … Guo, B. (2010). A SSR-based composite genetic linkage map for the cultivated peanut (*Arachis hypogaea* L.) genome. BMC Plant Biology, 10. https://doi.org/10.1186/1471-2229-10-17

Kottapalli, K.R., M.D. Burow, G. Burow, J. Burke, and N. Puppala. 2007. Molecular characterization of the U.S. peanut Mini-core collection using microsatellite markers. Crop Sci. 47(4): 1718–1727.

Krapovickas, A. and Rigoni, V.A., 1960. The nomenclature of the subspecies and varieties of A. *hypogaea* L. Revista de Investigaciones Agricoles, 14, pp.198-228.

Li, J. (2016). Association Mapping of SSR Markers to Tomato Spotted Wilt Virus Resistance in Cultivated Peanuts.

Liang, X., Chen, X., Hong, Y., Liu, H., Zhou, G., Li, S., & Guo, B. (2009). Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L .) and *Arachis* wild species. 9, 1–9. https://doi.org/10.1186/1471-2229-9-35

Liu, K., and S. V Muse. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21(9): 2128–9.

Mace, E.S., D.T. Phong, H.D. Upadhyaya, S. Chandra, and J.H. Crouch. 2006. SSR analysis of cultivated groundnut (*Arachis hypogaea* L.) germplasm resistant to rust and late leaf spot diseases. Euphytica 152(3): 317–330.

Macedo, S.E., M.C. Moretzsohn, S.C. M Leal-Bertioli, D.M. Alves, E.G. Gouvea, V.C. Azevedo, and D.J. Bertioli. 2012. Development and characterization of highly polymorphic long TC repeat microsatellite markers for genetic analysis of peanut. BMC Res. Notes 5(1): 86.

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): 1–11.

Mohsen, S. M., & Ammar, A. S. M. (2009). Total phenolic contents and antioxidant activity of corn tassel extracts. Food Chemistry, 112(3), 595–598. https://doi.org/10.1016/j.foodchem.2008.06.014

Moretzsohn, M.C., L. Leoi, K. Proite, P.M. Guimarães, 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(6): 1060–71.

Pandey, M.K., H.D. Upadhyaya, A. Rathore, V. Vadez, M.S. Sheshshayee, M. Sriswathi, M. Govil, A. Kumar, M.V.C. Gowda, S. Sharma, F. Hamidou, V.A. Kumar, P. Khera, R.S. Bhat, A.W. Khan, S. Singh, H. Li, E. Monyo, H.L. Nadaf, G. Mukri, S.A. Jackson, B. Guo, X. Liang, and R.K. Varshney. 2014. Genomewide association studies for 50 agronomic traits in peanut using the "reference set" comprising 300 genotypes from 48 countries of the semi-arid tropics of the world. PLoS One 9(8).

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

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular biology and evolution, p.mst197.

Tang, R., G. Gao, L. He, Z. Han, S. Shan, R. Zhong, C. Zhou, J. Jiang, Y. Li, and W. Zhuang. 2007. Genetic Diversity in Cultivated Groundnut Based on SSR Markers. J. Genet. Genomics 34(5): 449–459.

TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics 23(19): 2633–2635.

Vignal, A., D. Milan, M. SanCristobal, and A. Eggen. A review on SNP and other types of molecular markers and their use in animal genetics. Genet. Sel. Evol. 34(3): 275–305. 72

Wang, M.L., S. Sukumaran, N. a. Barkley, Z. Chen, C.Y. Chen, B. Guo, R.N. Pittman, H.T. Stalker, C.C. Holbrook, G. a. Pederson, and J. Yu. 2011. Population structure and marker- trait association analysis of the US peanut (*Arachis hypogaea* L.) Mini-core collection. Theor. Appl. Genet. 123(8): 1307–1317.

Yu, J., G. Pressoir, W.H. Briggs, I. Vroh Bi, M. Yamasaki, J.F. Doebley, M.D. McMullen, B.S. Gaut, D.M. Nielsen, J.B. Holland, S. Kresovich, and E.S. Buckler. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38(2): 203–8.

Zhang, J., Liang, S., Duan, J., Wang, J., Chen, S., Cheng, Z., … Li, Y. (2012). De novo assembly and Characterisation of the Transcriptome during seed development, and generation of genic-SSR markers in Peanut (*Arachis hypogaea* L.). BMC Genomics, 13(1). https://doi.org/10.1186/1471-2164-13-90

Table 3.1 The list of 133 SSR markers used in the study.

Marker	Marker name	Repeat motif	Reference
A2	Ad90K4	AT	Leal-Bertioli *et al.,*2009
A3	Ad90P7	TA	Leal-Bertioli *et al.,*(unpublished)
A4	Ad91O22	AG	Leal-Bertioli *et al.,*2009
A5	Ad92L5	AT	Leal-Bertioli *et al.,*2009
A8	Ad95D7	AT	Leal-Bertioli *et al.,*2009
A9	Ai119F10	TA	Leal-Bertioli *et al.,*2009
A10	Ai119G12	CAT/TTA	Leal-Bertioli *et al.,*2009
A12	Ai120L22	AT	Leal-Bertioli *et al.,*(unpublished)
B1	Ai121G16	TTA	Leal-Bertioli *et al.,*2009
B2	TC4F02	C	Moretzsohn *et al.,*(2005)
B4	TC13E05	TC	Macedo *et al.,*(2012)
B7	TC19B07	TC/CT	Macedo *et al.,*(2012)
B8	TC19B11	TC	Macedo *et al.,*(2012)
B10	TC20B05	GA/TA	Macedo *et al.,*(2012)
B11	TC20E08	GT/GA/TG	Macedo *et al.,*(2012)
B12	TC21C10	GA	Macedo *et al.,*(2012)
C1	TC21D06	CT	Macedo *et al.,*(2012)
C2	TC23C08	GA	Macedo *et al.,*(2012)
C3	TC23E04	GA	Macedo *et al.,*(2012)
C4	TC23F04	GA	Macedo *et al.,*(2012)
C5	TC23H10	GA	Macedo *et al.,*(2012)
C6	TC24B05	TC	Macedo *et al.,*(2012)
C7	TC24C06	TC	Macedo *et al.,*(2012)
C9	TC25G11	CT	Macedo *et al.,*(2012)
C10	TC27H12	AG	Macedo *et al.,*(2012)
C11	TC28B01	AG/GA	Macedo *et al.,*(2012)
D1	TC29H08	CT	Macedo *et al.,*(2012)
D2	TC30D04	GT/GA	Macedo *et al.,*(2012)
D3	TC31E08	CT/CA	Macedo *et al.,*(2012)
D4	TC31G11	AG	Macedo *et al.,*(2012)
D5	TC31H02	GA	Macedo *et al.,*(2012)
D6	TC31H06	CT	Macedo *et al.,*(2012)
D7	TC38D06	AG	Macedo *et al.,*(2012)
D8	TC39B04	CT	Macedo *et al.,*(2012)
D9	TC39F01	AG/GA	Macedo *et al.,*(2012)

Table 3.1 Continued.

Marker	Marker name	Repeat motif	Reference
D10	TC42A02	AG	Macedo *et al.,*(2012)
D11	TC41A10	GA	Macedo *et al.,*(2012)
D12	TC41C11	AG	Macedo *et al.,*(2012)
E2	TC42A05	AG	Macedo *et al.,*(2012)
E3	Ah11.1	TTA	Gimenes *et al.,*(2007)
E4	Ah20	TG	Gimenes *et al.,*(2007)
E6	Ah32	NA	NA
E7	Ah51	AG	Gimenes *et al.,*(2007)
E9	Ah282.1	CCA/AAG	Gimenes *et al.,*(2007)
E11	GM699	AAT	Budiman *et al.,*(2006)
F4	GM1135	TCC	Nagy *et al.,*(2009)
F5	GM1609	CT	Nagy *et al.,*(2009)
F6	GM1733	AT	Nagy *et al.,*(2009)
F9	GM1856	CT	Nagy *et al.,*(2009)
F10	GM1867	AG	Nagy *et al.,*(2009)
G1	GM1907	TTC	Nagy *et al.,*(2009)
G2	GM1979	GA	Nagy *et al.,*(2009)
G3	GM1991	TC	Nagy *et al.,*(2009)
G5	GM2308	CT	Nagy *et al.,*(2009)
G6	GM2350	GA	Nagy *et al.,*(2009)
G7	GM2602	GAA	Nagy *et al.,*(2009)
G8	GM2603	GAA	Nagy *et al.,*(2009)
G11	PM31	CT	He *et al.,*(2003)
G12	PM69	NA	He *et al.,*(2007)
H1	PM85	NA	He *et al.,*(2009)
H2	PM308	CT	He *et al.,*(2003)
H3	PM348	AG	He *et al.,*(2005)
H11	PM671	GGC	He *et al.,*(2006)
H12	PM721	GGC	He *et al.,*(2008)
a1	GM2638	TC	Nagy *et al.,*(2009)
a3	GM2690	AGA/TTG	Nagy *et al.,*(2009)
a5	GM2745	TA	Nagy *et al.,*(2009)
a6	GM2723	GAA	Nagy *et al.,*(2009)
a7	GM2724	TTC	Nagy *et al.,*(2009)
a10	GM2769	CTT	Nagy *et al.,*(2009)
a11	GM2779	AAT	Nagy *et al.,*(2009)

Table 3.1 Continued.

Marker	Marker name	Repeat motif	Reference
a12	GM2767	GAA	Nagy *et al.,*(2009)
b1	GM2774	AT	Nagy *et al.,*(2009)
b4	GM2791	TATGA	Nagy *et al.,*(2009)
b7	GM2788	TCT/GGT	Nagy *et al.,*(2009)
b10	GM2831	CAA	Nagy *et al.,*(2009)
b11	GM2845	CTT	Nagy *et al.,*(2009)
b12	GM2839	CA	Nagy *et al.,*(2009)
cc1	pPGPseq1B9	AG	Ferguson *et al.,*(2004)
cc2	pPGPseq2B9	TTA	Ferguson *et al.,*(2004)
cc3	pPGPseq2G4	TAA	Ferguson *et al.,*(2004)
cc4	pPGPseq3B8	TTA	Ferguson *et al.,*(2004)
cc5	pPGPseq3E10	AAT	Ferguson *et al.,*(2004)
cc6	pPGPseq5D5	GA	Ferguson *et al.,*(2004)
cc7	pPGPseq7G2	ATAG	Ferguson *et al.,*(2004)
cc9	pPGSseq9G5	CT	Ferguson *et al.,*(2004)
c11	pPGSseq10D4	CT	Ferguson *et al.,*(2004)
cc12	pPGSseq11G3	TTC	Ferguson *et al.,*(2004)
d1	pPGSseq13A7	ATA	Ferguson *et al.,*(2004)
d2	pPGSseq13A10	TTA	Ferguson *et al.,*(2004)
d3	pPGSseq13E9	AAT	Ferguson *et al.,*(2004)
d4	pPGSseq14H6	AC	Ferguson *et al.,*(2004)
d5	pPGSseq15C10	AAT	Ferguson *et al.,*(2004)
d6	pPGSseq16C6	TC	Ferguson *et al.,*(2004)
d7	pPGSseq16G8	AT	Ferguson *et al.,*(2004)
d8	pPGSseq17E1	TAT	Ferguson *et al.,*(2004)
d9	pPGSseq18A5	TA	Ferguson *et al.,*(2004)
d10	pPGSseq18C5	ATA	Ferguson *et al.,*(2004)
d12	pPGSseq18G1	TTG	Ferguson *et al.,*(2004)
e2	pPGSseq19A5	ATT	Ferguson *et al.,*(2004)
e3	pPGSseq19B12	TTA	Ferguson *et al.,*(2004)
e4	pPGSseq19D6	AAT	Ferguson *et al.,*(2004)
e5	pPGSseq19D9	TAT	Ferguson *et al.,*(2004)
e8	pPGSseq17E3	TCT	Ferguson *et al.,*(2004)
e9	pPGSseq19B1	AG	Ferguson *et al.,*(2004)
e10	pPGPseq2C11	TTA	Ferguson *et al.,*(2004)
e11	pPGPseq2D12B	AAT	Ferguson *et al.,*(2004)
f1	pPGPseq2F5	TTA	Ferguson *et al.,*(2004)

Table 3.1 Continued.

Marker	Marker name	Repeat motif	Reference
f2	pPGPseq3B5	AG	Ferguson *et al.,*(2004)
f3	TC02D06	AG	Moretzsohn *et al.,*(2005)
f4	TC03A12	TC	Moretzsohn *et al.,*(2005)
f5	TC03E02	CT/CA	Moretzsohn *et al.,*(2005)
f7	TC04F12	CT	Moretzsohn *et al.,*(2005)
f8	TC04G02	TC	Moretzsohn *et al.,*(2005)
f10	TC05A06	TC	Moretzsohn *et al.,*(2005)
f11	TC05D06	AG	Moretzsohn *et al.,*(2005)
f12	TC06E01	GA	Moretzsohn *et al.,*(2005)
g1	TC06H03	AG	Moretzsohn *et al.,*(2005)
g4	TC11B04	GA	Moretzsohn *et al.,*(2005)
g5	TC2A02	CT	Moretzsohn *et al.,*(2005)
g6	TC3E05	GA	Moretzsohn *et al.,*(2005)
g7	gi4925	ATT	Moretzsohn *et al.,*(2005)
g8	RI1F06	ATA/ATT/GA	Moretzsohn *et al.,*(2005)
g9	AC2C05	TG	Moretzsohn *et al.,*(2005)
g10	Ah3	GA/AG/GT	Gimenes *et al.,*(2007)
g11	Ah26	CT	Gimenes *et al.,*(2007)
g12	Ah193	AAC/GA	Moretzsohn *et al.,*(2004)
h2	XIP297	TC	Cuc *et al.,*(2008)
h6	Lec1	AT	Hopkins *et al.,*(1999)
h7	XIP105	CT	Cuc *et al.,*(2008)
h8	XIP108	TC	Cuc *et al.,*(2008)
h9	XIP123	GA	Cuc *et al.,*(2008)
h10	XIP136	CT	Cuc *et al.,*(2008)

Table 3.2 Four subpopulations stratified by STRUCTURE analysis for 104 genotypes from the U.S. peanut Mini-core collection.

G1 (25)	G2 (18)	G3 (25)	G4 (36)
PI259617	PI152146	PI158854	PI290566
PI262038	PI200441	PI240560	PI343398
PI337406	PI259836	PI268586	PI159786
PI356004	PI290620	PI268696	PI162655
PI429420	PI295730	PI268755	PI162857
PI471954	PI313129	PI268806	PI196622
PI482189	PI339960	PI338338	PI196635
PI475918	PI478850	PI155107	PI259658
PI493329	PI502040	PI270998	PI259851
PI493356	PI475863	PI271019	PI157542
PI493547	PI476025	PI288146	PI268868
PI493581	PI502111	PI270786	PI268996
PI493631	PI502120	PI270905	PI274193
PI493693	PI270907	PI337293	PI290536
PI493717	PI494795	PI337399	PI290594
PI493729	PI497639	PI461434	PI292950
PI493880	PI497318	PI481795	PI295250
PI493938	PI497395	PI482120	PI295309
PI325943		PI504614	PI296550
PI331297		PI290560	PI296558
PI343384		PI407667	PI298854
PI471952		PI478819	PI319768
PI403813		PI476432	PI323268
PI497517		PI494018	PI331314
PI461427		PI494034	PI355268
			PI355271
			PI370331
			PI372271
			PI372305
			PI399581
			PI442768
			PI476636
			PI496401
			PI496448
			PI288210
			PI371521

Table 3.3 Genetic distances within four subpopulations detected by STRUCTURE analysis.

Group	G1	G2	G3	G4
G1	0.00	0.19	0.18	0.24
G2	0.37	0.00	0.17	0.16
G3	0.36	0.35	0.00	0.21
G4	0.43	0.34	0.40	0.00

The top diagonal is Nei's minimum distance and the bottom diagonal is pairwise F_{st}.

Table 3.4 Significantly SSRs associated identified by 4 models for sensory and quality attributes in peanut.

Traits		Marker	Q	PCA	Q + K	P + K	P-value	R^2 (%)
Fatty Acid	Raw Oil Content (% wet wt)	GM2690	√	√	√	√	0.0025	9.19
		GM2774	√	√	√	√	0.0015	10.19
		GM2791	√	√	√	√	0.0021	9.5
Tocopherols	Alpha	GM1609	√	√	√	√	<0.0001	15.12
		Ah32	√	√	√	√	0.0007	10.1
	Beta	pPGPseq2C11	√	√	√	√	0.0004	11.26
	Delta	XIP297	√	√	√	√	<0.0001	16.03
	Gamma	Ah3	√	√	√	√	0.0002	14.27
	Total	g10	√	√	√	√	0.0004	13.22
Sugars	Fructose	GM2690	√	√	√	√	0.002	9.4
	Glucose	GM2690	√	√	√	√	0.0004	10.76
	Sucrose	GM1609	√	√	√	√	<0.0001	12.9
Sensory Flavors	Fruity	pPGPseq5D5	√	√	√	√	0.0002	14.14
	Fermented	GM2745	√	√	√	√	<0.0001	12.9
	Sweet Aromatic	GM2723	√	√	√	√	<0.0001	12.9
	Raw Beany	GM1867	√	√	√	√	<0.0001	12.9

a: '√' means the marker is identified in the model.
b: The Marker_R^2 based on PCA model.

Table 3.5 Significantly SSRs associated identified by four algorithms for flavors profile.

		Q MODEL		PCA MODEL		Q + K MODEL		P + K MODEL	
Traits	**Maker**	**P value**	**R2 (%)**	**P value**	**R2 (%)**	**P value**	**R2 (%)**	**P value**	**R2 (%)**
	f8	0.002	8.53			0.005	7.48		
	A3	0.026	4.67			0.010	6.28		
	E3	0.025	4.75			0.010	6.19		
	a3	0.011	6.11			0.033	4.30		
	D11	0.025	4.77						
	E6	0.036	4.17						
Bitter	H2	0.049	3.68						
	B4	0.021	4.91	0.007	7.61	0.017	5.17		
	d5	0.034	4.17			0.021	4.89		
	a10	0.029	4.39			0.038	3.95		
	D9	0.026	4.56	0.003	9.45	0.038	3.95	0.008	7.07
	a3	0.039	3.94			0.041	3.85		
	b10	0.030	4.35			0.047	3.64		
	f8	0.029	4.40			0.047	3.64		
	F10	0.036	4.06						
	e9			0.012	6.65				
	f10			0.033	4.72				
DR	G1			0.050	3.97				
	cc6	0.000	12.95	0.000	13.22	0.000	13.97	0.000	14.14
	a5	0.000	11.40	0.001	11.73	0.001	9.73	0.002	10.05
	D12	0.013	5.73	0.012	6.30	0.008	6.59	0.007	7.39
	h9	0.010	6.08	0.018	5.60	0.011	6.09	0.024	5.19
	E7	0.045	3.74						
FF	cc7	0.048	3.66						
	d12	0.000	34.66	0.000	37.86	0.000	32.07	0.000	37.52
	H1	0.002	9.73	0.002	10.75	0.003	8.22	0.002	10.53
	cc2	0.013	6.12			0.008	6.70	0.040	4.47
	g10	0.007	7.22			0.014	5.80		
	e3	0.005	7.59	0.009	7.35	0.037	4.17	0.010	7.09
	h8	0.023	5.15			0.046	3.81		
	H2	0.012	6.23	0.037	4.58			0.034	4.75
	cc5	0.034	4.48						
	h10	0.036	4.38						
Painty	F6			0.042	4.32			0.045	4.25
	a6	0.002	8.47	0.001	10.68	0.001	10.11	0.001	12.49
	f2	0.006	6.72	0.007	7.26	0.002	9.01	0.005	8.18
	b1	0.011	5.70	0.005	7.77	0.002	8.60	0.002	10.28
	G11	0.036	3.93			0.009	6.35	0.036	4.51
	G1	0.043	3.65			0.036	4.14		
	h7	0.033	4.05	0.027	4.80			0.049	3.96
SA	B8	0.038	3.83	0.030	4.63			0.039	4.39
	f8	0.003	8.09			0.005	6.91		
	F10	0.001	9.33	0.022	5.44	0.008	6.39	0.027	5.08
	A3	0.035	4.14			0.021	4.85		
	e5	0.048	3.65			0.037	3.98		
	e9			0.015	6.23	0.050	3.53	0.019	5.67
	a3	0.036	4.08						
	b12	0.046	3.71						
RB	G2			0.010	6.97			0.011	6.76

	F5			0.020	5.64			0.018	5.84
	D9			0.022	5.44			0.038	4.41
	cc5							0.042	4.24
	F5					0.008	6.49	0.042	4.07
	C10					0.045	3.81		
	d7	0.018	5.30						
	f8	0.030	4.51						
	G3	0.034	4.31	0.048	4.02				
Sweet	D8	0.049	3.71					0.028	4.79
	a6	0.010	6.16	0.013	6.32	0.005	7.30	0.009	7.13
	f2	0.013	5.69	0.010	6.94	0.007	6.85	0.009	7.23
	b1	0.016	5.40	0.043	4.17	0.011	6.16		
	a10	0.019	5.08			0.014	5.74		
	C7	0.037	4.06			0.014	5.74	0.041	4.30
	A3	0.043	3.81	0.013	6.32	0.018	5.36	0.011	6.71
	G11	0.038	4.01			0.019	5.20		
	b10	0.032	4.26			0.021	5.10		
	f7	0.043	3.81			0.032	4.42	0.046	4.08
	G1	0.048	3.64			0.036	4.21		
	E4	0.024	4.74	0.045	4.08	0.038	4.14		
	h7	0.025	4.66	0.010	6.88	0.039	4.09	0.014	6.26
RP	D12			0.012	6.57			0.005	8.26
	d5	0.017	5.59	0.027	5.15	0.017	5.62	0.028	5.15
	a11	0.025	4.99			0.020	5.32		
	A9	0.015	5.82			0.024	5.08		
	cc3	0.029	4.70			0.042	4.12		
	D10	0.049	3.86			0.043	4.07		
	B8	0.039	4.22			0.045	4.00		
	f4	0.048	3.89			0.047	3.95		
	A3	0.019	5.41						
Spice	G6	0.047	3.91						
	f8	0.008	6.83			0.011	6.25		
	d1	0.010	6.45	0.009	7.23	0.014	5.84	0.034	4.39
	B7	0.022	5.10	0.049	3.98	0.021	5.15		
	B2	0.037	4.21	0.015	6.14	0.030	4.59	0.007	7.25
	D9	0.043	3.99	0.029	4.92	0.044	3.97	0.030	4.64
	G1	0.048	3.80			0.045	3.93		
	C4	0.019	5.31	0.024	5.25			0.036	4.29
	a12	0.044	3.96						
WHS	cc9	0.049	3.78						

Table 3.6 Significantly SSRs associated identified by four algorithms for sugars profile.

		Q MODEL		PCA MODEL		Q + K MODEL		P + K MODEL	
Traits	**Maker**	**P value**	**R2 (%)**	**P value**	**R2 (%)**	**P value**	**R2 (%)**	**P value**	**R2 (%)**
	a3	0.001	8.73	0.006	7.29	0.000	9.16	0.002	9.40
	g7	0.003	6.62			0.002	7.11		
	F10	0.021	4.03	0.027	4.59	0.006	5.79	0.009	6.68
	h7	0.001	7.70	0.002	8.87	0.007	5.46	0.004	7.93
	cc11			0.030	4.43	0.007	5.45	0.012	6.15
	h6	0.011	4.89	0.018	5.27	0.011	4.92	0.013	6.00
	g1	0.028	3.68	0.027	4.58	0.013	4.75	0.025	4.83
	a6	0.038	3.30			0.018	4.29		
	b7	0.033	3.49			0.024	3.90		
	A3					0.042	3.19		
	a11	0.008	5.37	0.014	5.67	0.044	3.12	0.033	4.35
	cc5	0.005	5.81	0.003	8.34			0.022	5.08
	B10	0.011	4.93	0.041	3.92				
	G12	0.017	4.30						
	F5	0.018	4.28						
	g10			0.031	4.36			0.016	5.57
Fructose	A8			0.045	3.78			0.050	3.69
	e9					0.018	3.69	0.026	4.36
	d5					0.029	3.11	0.037	3.83
	E6			0.049	3.47	0.031	3.05	0.038	3.80
	f3	0.012	4.06	0.024	4.59	0.036	2.90	0.024	4.50
	b4				0.00	0.045	2.65		
	B4					0.049	2.56		
	C1	0.018	3.61	0.024	4.62				
	B1	0.031	3.00	0.037	3.92				
Raffinose	f7	0.046	2.58		0.00				
	a3	0.000	10.76	0.002	9.45	0.002	8.80	0.004	8.74
	F5	0.000	12.29	0.038	4.18	0.003	7.99	0.022	5.34
	cc11					0.003	7.72	0.007	7.31
	A3					0.007	6.48		
	C6	0.011	5.68	0.049	3.73	0.011	5.84	0.039	4.28
	g7	0.008	6.16			0.015	5.38		
	e2	0.004	7.40	0.049	3.75	0.015	5.36		
	B11	0.028	4.33			0.023	4.68		
	a5					0.033	4.16		
	h6	0.023	4.62	0.043	3.98	0.035	4.07	0.039	4.29
	g5					0.042	3.79	0.045	4.05
	b7	0.046	3.58			0.044	3.69		
	F4	0.002	8.21	0.008	7.01				
	b12	0.007	6.46	0.034	4.38				
	h7	0.014	5.34						
	G12	0.020	4.80						
	f10	0.021	4.75						
	e3	0.034	4.04						
	G1	0.037	3.91						
Glucose	cc7	0.050	3.46						
	G5	0.007	7.01	0.009	7.16	0.007	7.09	0.024	5.13
	h6	0.020	5.34	0.015	6.07	0.013	6.00	0.027	4.95
	e5	0.008	6.82	0.046	4.04	0.013	5.99	0.017	5.81
Inositol	G3	0.032	4.55		0.00	0.015	5.76		0.00

	D2	0.022	5.12	0.035	4.53	0.023	5.03	0.038	4.31
	B11	0.021	5.27	0.022	5.42	0.034	4.39		0.00
	H2	0.041	4.11	0.040	4.31		0.00	0.038	4.34
	A4		0.00	0.047	4.04		0.00	0.031	4.67
	b10		0.00	0.032	4.73		0.00	0.042	4.15
	F5	0.000	12.90	0.002	9.73	0.000	9.78	0.004	8.16
	cc12					0.001	8.64	0.010	6.51
	e2	0.001	8.52	0.018	5.29	0.003	7.08	0.013	5.95
	b1	0.019	4.40	0.021	5.02	0.003	6.64	0.010	6.49
	E6	0.020	4.32	0.010	6.37	0.007	5.79	0.004	7.92
	d7	0.006	5.90	0.006	7.30	0.012	4.97	0.008	6.73
	e5	0.013	4.89			0.017	4.48		
	b4	0.030	3.75	0.022	4.96	0.020	4.32	0.016	5.58
	d5	0.009	5.37	0.033	4.32	0.030	3.72		
	cc7	0.017	4.55	0.047	3.72	0.035	3.55		
	g10					0.040	3.37	0.037	4.19
	cc5					0.042	3.28		
	G12	0.006	5.95			0.044	3.25	0.019	5.25
	b7					0.045	3.21		
	B10	0.010	5.25			0.049	3.08		
	e3	0.012	5.03						
	h7	0.019	4.38						
	g7	0.021	4.23						
	f1	0.026	3.96						
	h9	0.043	3.29	0.031	4.39			0.047	3.77
Sucrose	d9	0.050	3.10						

Table 3.7 Significantly SSRs associated identified by four algorithms for fatty acids profile.

		Q MODEL		PCA MODEL		Q + K MODEL		P + K MODEL	
Traits	**Maker**	**P value**	**R2 (%)**	**P value**	**R2 (%)**	**P value**	**R2 (%)**	**P value**	**R2 (%)**
	a3	0.008	6.12	0.002	9.19	0.001	8.83	0.005	7.72
	b1	0.003	7.49	0.001	10.19	0.003	7.93	0.002	9.82
	g5	0.010	5.76	0.023	5.11	0.003	7.73	0.018	5.47
	cc12	0.018	4.84			0.005	6.93	0.048	3.77
	b4	0.009	5.89	0.002	9.50	0.006	6.67	0.004	8.27
	g10	0.004	6.94	0.009	6.71	0.006	6.66	0.006	7.40
	D1			0.046	3.88	0.015	5.33		
Raw Oil	E6	0.016	5.04	0.048	3.82	0.015	5.30	0.025	4.86
Content	F5	0.011	5.64					0.029	4.62
(%wetwt)	b7	0.047	3.45						
	B8	0.004	4.90	0.012	5.72	0.006	4.68	0.017	5.17
	G1	0.006	4.61	0.017	5.10	0.010	4.16	0.030	4.26
	A8	0.035	2.73			0.022	3.27		
	H3	0.049	2.39						
IV	cc1			0.011	5.79			0.013	5.55
	h2	0.011	5.72	0.042	4.09	0.008	6.14	0.034	4.43
	d9	0.029	4.28			0.012	5.59	0.041	4.10
	B12	0.009	5.99	0.048	3.88	0.017	5.04	0.046	3.90
	a11	0.015	5.31			0.018	4.98		
	E3	0.024	4.55	0.049	3.83	0.032	4.10		
	cc2	0.040	3.81	0.036	4.36	0.037	3.86	0.032	4.50
	f10	0.029	4.29	0.026	4.91	0.043	3.66	0.041	4.08
	D10	0.044	3.67						
	h8	0.047	3.56	0.047	3.90				
LCS	H1	0.048	3.53						
	D1	0.032	2.20			0.026	2.34		
	B8	0.019	2.60			0.027	2.31		
	H2	0.049	1.86			0.039	2.03		
	G1	0.022	2.49	0.028	3.99	0.044	1.94		
	D2	0.041	1.99						
	cc1			0.014	5.09			0.018	4.61
	b4			0.049	3.19			0.038	3.54
	F9			0.039	3.50				
O/l	d2			0.047	3.26				
	D9	0.010	5.75			0.003	7.77	0.030	4.73
	B8	0.003	7.38	0.003	8.86	0.005	7.03	0.005	8.10
	cc4	0.031	4.03			0.018	5.07		
	G1	0.017	4.93	0.048	3.79	0.018	5.03		
	d1	0.032	3.96			0.037	3.93		
	D2	0.011	5.59	0.035	4.30				
	cc1			0.018	5.48			0.021	5.38
	a11			0.025	4.92			0.029	4.80
P/S	b4			0.036	4.26			0.032	4.64
	G11	0.010	2.63			0.004	3.04		
	F9	0.016	2.32			0.012	2.38		
	A3	0.034	1.81			0.016	2.19		
	D1	0.041	1.68						
	D4	0.043	1.65						
Sat(%)	h8			0.030	3.79			0.023	3.96

Table 3.8 Significantly SSRs associated identified by four algorithms for tocopherols profile.

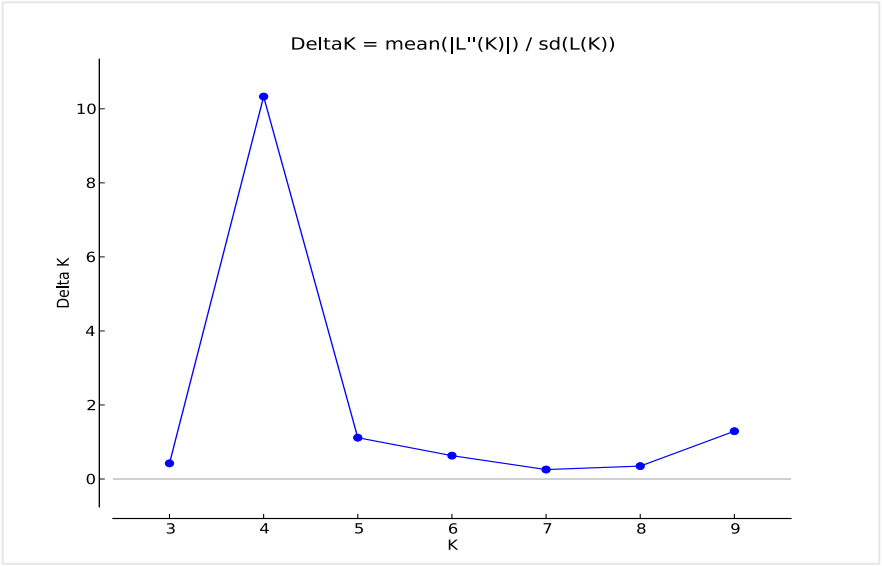
		Q MODEL		PCA MODEL		Q + K MODEL		P + K MODEL	
Traits	**Maker**	**P value**	**R2 (%)**	**P value**	**R2 (%)**	**P value**	**R2 (%)**	**P value**	**R2 (%)**
	B11	0.053	1.55	0.007	6.53	0.032	4.64	0.047	3.90
	b7	0.071	0.47	0.003	7.94	0.034	4.52		
	cc1	0.050	1.84	0.029	4.33				
	cc12	0.035	4.89	0.003	7.94	0.045	4.02		
	e2	0.070	0.52	0.011	5.79				
	e3	0.078	0.30	0.017	5.11	0.048	3.91	0.025	5.05
	E3	0.055	1.38	0.022	4.75				
	E6	0.080	0.26	0.001	10.06	0.007	7.39	0.008	7.13
	F5	0.151	0.00	0.000	14.70	0.002	10.23	0.002	9.60
	f8	0.056	1.27	0.033	4.10				
	G1	0.039	3.96	0.041	3.80				
	g10	0.041	3.33	0.027	4.44	0.020	5.41	0.022	5.27
	g6	0.075	0.36	0.028	4.36			0.040	4.18
	h2	0.053	1.51	0.012	5.71	0.033	4.57	0.038	4.28
	H2	0.040	3.66	0.017	5.16				
	b4			0.049	3.53	0.002	9.76	0.002	9.37
Alpha	e5					0.030	4.75	0.027	4.88
	e10	0.000	11.26	0.000	12.55	0.004	8.60	0.004	8.08
	h9	0.012	5.91	0.012	6.02	0.039	4.22	0.047	3.81
	D6	0.013	5.76	0.008	6.72				
	d8	0.021	4.99	0.017	5.37				
	b4	0.026	4.69	0.023	4.91	0.007	7.36	0.005	7.77
	f1	0.029	4.50	0.026	4.75	0.038	4.27	0.027	4.77
	h7	0.029	4.50	0.021	5.09				
	b7	0.031	4.40	0.017	5.38				
	a1	0.031	4.38	0.039	4.08				
	f7					0.010	6.68	0.009	6.66
Beta	e9					0.011	6.42	0.012	6.09
	h2	0.000	15.37	0.000	16.03	0.006	7.80	0.001	11.81
	h9	0.000	14.08	0.000	15.93	0.014	6.12		
	D8	0.000	12.95	0.001	10.35	0.033	4.57		
	a10	0.002	8.91	0.002	9.24	0.003	8.96		
	b1	0.004	7.63	0.002	9.16	0.031	4.71	0.018	5.56
	B11	0.014	5.77	0.028	4.73	0.025	5.11	0.028	4.73
	B7	0.019	5.27	0.042	4.07	0.030	4.77	0.030	4.64
	h7	0.019	5.24	0.011	6.24				
	a12	0.028	4.66	0.014	5.87				
	D6	0.032	4.45	0.041	4.10				
Delta	g9	0.047	3.81	0.034	4.41				
	g10	0.000	13.65	0.000	11.89	0.000	13.40	0.000	14.27
	E3	0.001	9.60	0.004	7.51	0.002	9.59	0.002	9.59
	E4	0.003	7.95	0.006	6.93	0.017	5.78	0.014	5.98
	h9	0.003	7.84	0.002	8.71	0.000	13.46	0.001	12.17
	D3	0.003	7.70	0.004	7.39	0.008	7.30	0.007	7.23
	D8	0.003	7.58	0.026	4.53			0.024	5.02
	F9	0.008	6.35					0.017	5.69
	h2	0.015	5.29	0.015	5.38	0.049	3.91	0.044	3.99
	D6	0.021	4.77	0.029	4.39	0.034	4.54	0.033	4.48
Gamma	D12	0.028	4.35	0.009	6.27				

	b4	0.050	3.49	0.033	4.19				
	F9			0.006	6.98	0.012	6.40		
	d3			0.015	5.44	0.037	4.41		
	a11			0.017	5.17	0.020	5.46	0.049	3.81
	f12			0.041	3.84	0.045	4.04		
	e2			0.045	3.71	0.038	4.35		
	cc9					0.029	4.80	0.041	4.11
	g10	0.000	13.24	0.000	12.52	0.000	13.22	0.000	12.84
	E3	0.000	12.47	0.001	10.31	0.018	5.65	0.021	5.33
	F5	0.000	11.31	0.003	7.94	0.018	5.73	0.010	6.71
	b7	0.001	10.40	0.001	10.21	0.010	6.82	0.013	6.14
	G12	0.007	6.70	0.020	5.02				
	D6	0.009	6.31	0.005	7.12	0.011	6.60	0.013	6.20
	b4	0.014	5.65	0.006	6.97	0.003	8.88	0.003	8.79
	E6	0.014	5.62	0.001	9.26	0.006	7.89	0.013	6.18
	G1	0.041	3.90	0.042	3.88				
	F9	0.045	3.75	0.041	3.92				
	d6			0.007	6.79	0.025	5.07		
Total	h9					0.037	4.36	0.038	4.29

Table 3.9 Summary of SSRs from Association Mapping identified within 3 Mb window of significant SNPs by GWAS analysis.

Trait	Linkage Group	SNP ID	Position (bp)	- $\log_{10}$ (*P* value)	SSR Marker	Position (bp)	R2 (%)	Distance (Mb)
	B03	AX-147243168	3566707	4.17	GM1609	3250954	6.49	0.32
	A01	AX-176794174	104118984	3.35	TC27H12	105295222	3.81	1.18
Sweet	A04	AX-176814619	39598280	3.12	TC39B04	40709402	4.79	1.11
	B03	AX-147243168	3566707	4.42	GM1609	3250954	12.90	0.32
	A02	AX-176809854	68712646	3.21	pPGSseq11G3	68230065	8.64	0.48
sucrose	B04	AX-176793698	84313635	4.44	pPGSseq19A5	85119881	8.52	0.81
	B04	AX-176813712	9074346	4.81	GM1867	9159865	6.68	0.09
	A08	AX-176809466	4403235	3.27	GM2788	4281833	3.90	0.12
	A06	AX-147226746	110928698	4.86	GM2779	113077149	5.67	2.15
Fructose	A10	AX-176806924	7788634	3.04	pPGPseq3E10	9071539	8.34	1.28
	B01	AX-176815914	134636409	3.19	TC30D04	135364360	5.12	0.73
	A01	AX-176818319	101370991	3.13	TC20E08	102123627	5.42	0.75
Inositol	B06	AX-176822690	15756056	3.23	GM2831	14368158	4.73	1.39
ROC	A03	AX-176814232	2814162	3.84	TC2A02	2270994	7.73	0.54
	B04	AX-176791423	131792362	3.10	GM1907	133168886	5.03	1.38
P/S	A02	AX-176804911	22662900	4.22	pPGPseq1B9	22408511	5.48	0.25
	B03	AX-176820466	4311972	3.75	TC29H08	4478383	2.34	0.17
	A02	AX-176804911	22662900	3.68	pPGPseq1B9	22408511	5.09	0.25
O/L	B04	AX-176800740	118083731	3.38	pPGSseq13A10	115972720	3.26	2.11
	B09	AX-177643501	10735224	3.29	XIP297	10868851	6.14	0.13
	B09	AX-177643084	49917116	3.10	TC05A06	52566323	4.91	2.65
LCS	A04	AX-147221053	117995189	3.58	XIP108	117391812	3.90	0.60
	B10	AX-177637601	103487554	3.16	TC19B11	101191272	5.72	2.30
IV	A02	AX-176804911	22662900	4.60	pPGPseq1B9	22408511	5.79	0.25
α-Toco	B06	AX-176808373	104214231	5.23	pPGSseq19A5	101914706	5.79	2.30

	B04	AX-147248776	130709725	3.86	GM1907	133168886	3.96	2.46
	B08	AX-177638150	108958569	3.73	XIP123	106387141	6.02	2.57
β-Toco	A04	AX-147220926	115648648	3.47	XIP105	117579778	5.09	1.93
	B08	AX-147257537	7765731	5.67	Ah20	7879276	7.95	0.11
	B09	AX-176791660	143735714	4.94	GM1856	143256842	6.35	0.48
	A06	AX-176805357	94179951	4.18	TC31H06	94629862	4.77	0.45
	B06	AX-176808373	104214231	4.14	pPGSseq19A5	101914706	4.35	2.30
γ-Toco	A05	AX-176822894	32830345	3.56	pPGSseq9G5	32425313	4.80	0.41
	B08	AX-177638150	108958569	3.92	XIP123	106387141	15.93	2.57
δ-Toco	B02	AX-176791511	30172328	4.68	AC2C05	28006100	4.41	2.17
	A06	AX-176805357	94179951	5.74	TC31H06	94629862	7.12	0.45
Total Toco	B09	AX-176791660	143735714	4.50	GM1856	143256842	3.92	0.48



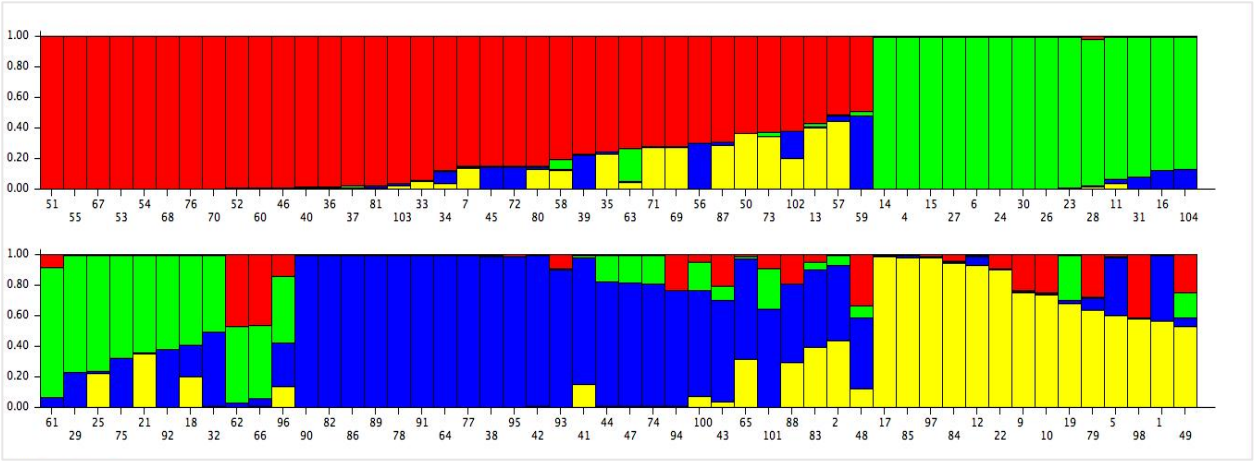


Fig. 3.1 Population structure analysis. A) Magnitude of Δ K by STRUCTURE analysis of the U.S. peanut Mini-core collection. B) Population structure analysis. The y-axis stands for the subgroup membership, and x-axis stands for the genotypes. G1-G4 represents for four subpopulations.

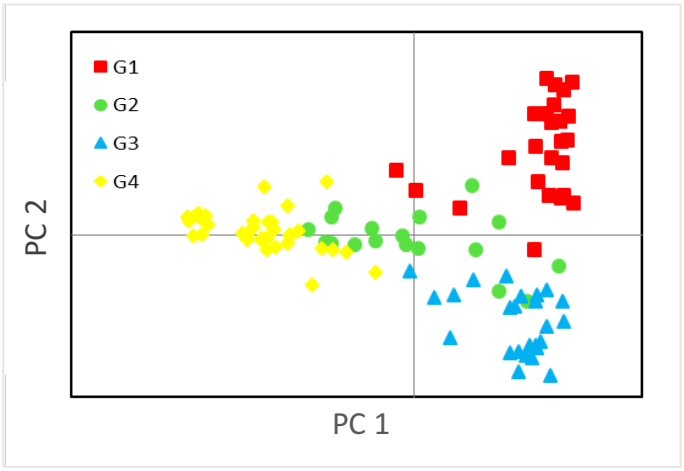


Fig. 3.2 Principle component analysis (PCA) plot for the U.S. peanut Mini-core collection. The branch colors are consistent with the colors of the STRUCTURE clusters.

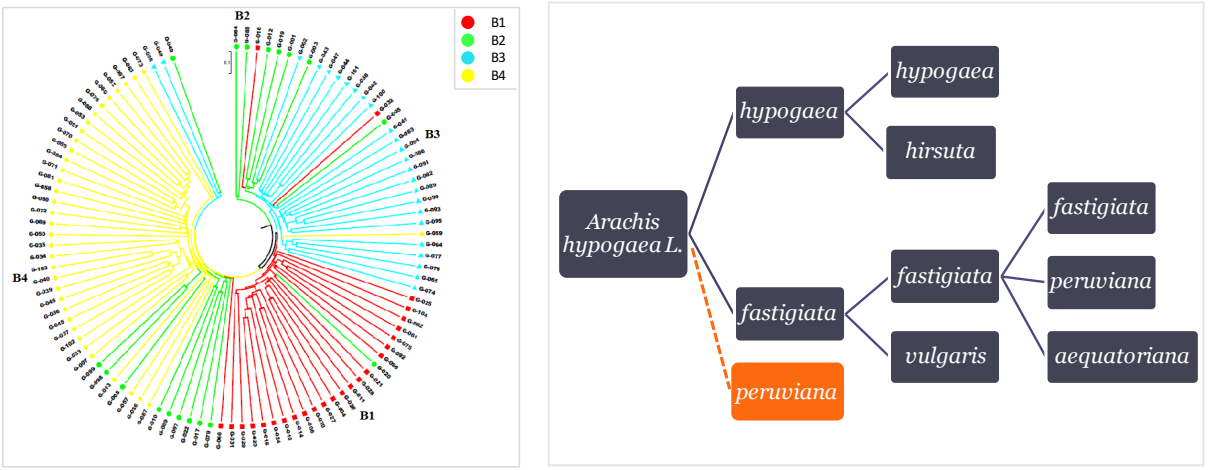


Fig. 3.3 A) UPGMA tree plot based on Chord distance for the U.S. peanut Mini-core collection. The tree branch colors are consistent with the colors of the structure clusters. B) Traditional and new constructed hierarchy of peanut variety.

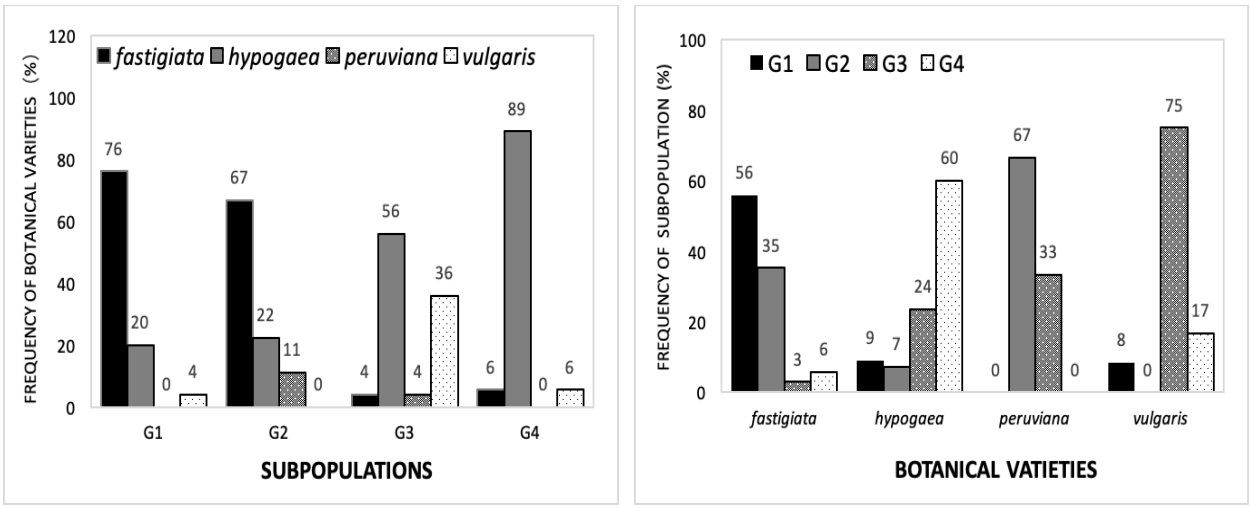


Fig. 3.4 A) Frequency of botanical variety within each subpopulation. B) Frequency of each subpopulation within each botanical variety.

Chapter Four

Next-Generation RNA-Sequencing Identified Genomic Regions Involving in High Oleic Acid Accumulation during Pod Development Stages in Peanut

4.1 Abstract

As obtaining desirable flavor and quality attributes becomes an indispensable breeding objective for peanut, high oleic acid varieties had been proven to have higher burnt peanut aroma and burnt peanut flavor compared to regular oleic peanuts. The objective of this study was to identify the genes involved in oleic acid metabolic networks in different peanut genotypes and growth stages. To accomplish this goal, four genotypes (AABB, AAbb, aaBB, and aabb) in four pod development stages (yellow, orange, brown, and black stages) were examined. Whole-transcriptome sequencing analysis identified 526 genes differentially expressed in AABB, 923 genes in AAbb, 2,286 genes in aaBB, and 758 genes in aabb, while 369 genes differentially expressed in Yellow stage, 1,406 genes in Orange stage, 732 genes in Brown stage and 1519 genes in the Black stage. Functional analysis of the shared DEGs showed that most genes are those processes involved in metabolic pathways, biosynthesis of secondary metabolites, fatty acids, lipids biosynthesis, carbohydrates metabolism. Besides, 31 DEGs were found to locate within 3 Mb of significant SNPs associated with O/L ratio, including three DEGs in Chromosome B04, which were found locating 0.08 Mb from AX-176800740 and locating 2.11 Mb from SSR marker pPGSseq13A10. It would help to understand the mechanisms of the oleic acid metabolic network during different peanut pod growth stages.

4.2 Introduction

Cultivated peanut (*Arachis hypogaea* L.) is an important food crop with many health benefits of their consumption appreciated by consumers worldwide. Ascribable to the adaptation of the high oleic acid trait into several new cultivars, breeders could accumulate the sensory and quality attributes related to cultivated peanuts much more efficient (Talcott et al., 2005; Vassiliou et al., 2009). Peanuts with high oleic acid contents were determined to have improved stability against lipid oxidation that could lead to adverse flavors (Talcott et al., 2005). Also, high oleic acid varieties had been proven to not only have higher burnt peanut aroma and burnt peanut flavor compared to regular oleic peanuts but also is intimately bound up the health of human beings. The enzymes, ahFAD2A and ahFAD2B, are coded by two essential homologous genes. The double mutant ahfad2a/ahfad2b has been identified and proved to have a high O/L ratio (Chu et al., 2011). How to sustain and even increase peanut high oleic acid content to meet growing needs for sensory and quality attributes is a significant challenge that the peanut industry faces. Developing high oleic acid varieties is a crucial objective for many of peanut breeding programs (Zhao et al., 2018). However, the inefficient genetic methodology and a limited amount of genetic information on *Arachis* species have impeded the process of molecular mechanism study in peanut cultivars.

The lasted developments in large genome sequencing technologies have enhanced opportunities to develop and disseminate genetic resources at a lower expense (Chopra et al., 2014). The advent of RNA-sequencing (RNA-Seq), a rapid technique for genome-wide gene expression analysis, provides powerful alternatives to facilitate the identification of different functional genes in a more efficient manner (Lu et al., 2014). Subsequently, gene structures and expression profiles in many crops, including wheat, corn, soybean, Brassica rapa, and peanut, were determined with RNA-Seq technology (Strauß et al., 2012; Li et al., 2014; Bhardwaj et al., 2015; Brasileiro et al., 2015; Zhao et al., 2018; Ruan et al., 2018; Lang et al., 2019). As the genome sequence of cultivated peanut

released (Bertioli et al., 2019), a more accurate transcriptome assembly becomes possible as it could be mapping to the reference genome. Thus, the objective of the present study was to discover oleic acid-related genes in comparing different genotypes, including high oleic acid line and normal oleic acid line under different growth stages by mapping RNA-Seq data to the cultivated peanut reference genome.

The data developed in this work could provide comprehensive insight into molecular mechanisms that underlie oleic acid metabolic network and resource to further molecular research in peanut.

4.3 Materials and Methods

4.3.1 Materials and experimental design

Four RIL peanut genotypes, AABB (normal oleic acid), AAbb (medium oleic acid), aaBB (medium oleic acid), and aabb (high oleic acid) were used in this study. The different maturity seed samples were collected in four pod growth stages, representing Yellow, Orange, Brown, and Black. All seeds were planted in a single-row (15 x 120 cm) at a rate of 6 seeds m-1 under rainout shelters at the USDA-ARS National Peanut Research Laboratory in Dawson.

4.3.2 RNA isolation, library construction, and Illumina sequencing

Different growth stages of pods were collected from each genotype with three replications. The samples were flash-frozen by liquid nitrogen and stored at -80 °C until RNA extraction. Each RNA extraction samples were pooled for three biological replications. Approximately 0.2 g pooled samples were ground in liquid nitrogen for RNA extraction by using a modified CTAB method (Yin et al., 2011) and purified by using a Direct-Zol RNA MiniPrep Kit (Zymo Research, Irvine, CA, USA). The purity and integrity of RNA were analyzed by NanoDrop ND-1000 UV spectrophotometer (Thermo Scientific, Wilmington, USA) and Agilent 2100 Bioanalyzer (Agilent,

USA). A total of 48 libraries (4 genotypes × 4 growth stages × 3 replicates) were constructed and subsequently sequenced using Illumina HiSeq 4000 at the Beijing Genomics Institute (BGI).

4.3.3 Quality control, alignment and genome-guided assembly

The raw reads from RNA-sequencing were pruned with Trimmomatic (Bolger et al., 2014). We removed the adaptor sequences, ambiguous 'N' nucleotides, and low-quality reads from the raw data in order to obtain the cleaned reads. The read quality was assessed by FastQC (Andrews, 2010) before and after trimming. High-quality clean data was subjected to the downstream analyses. The RNA-seq data analysis pipeline strictly followed the protocol from Trapnell et al. (2012). Each sample was mapped to the reference genome by Tophat2 (Kim, 2013), with all the parameters set to default. The cultivated peanut genome and the annotation file (Bertoli et al., 2019) were used as a reference for alignment. The alignment files of the 16 samples from Tophat2 were input into Cufflinks (Trapnell et al., 2012) for transcripts reconstruction. To identify the novel transcript sequences, all the assemblies were compared with the reference annotation by using Cuffcompare. Novel transcript sequences were then compared to the 'nr' database at NCBI by BLASTX to achieve functional gene annotation.

4.3.4 Identification of differentially expressed genes

The gene expression levels were represented by the expected number of Fragments Per Kilobase of transcript sequence per Millions of base pairs sequenced (FPKM), which was calculated based on the length of the gene and reads count mapped to the genes. The DEGs analysis was performed by using Cuffdiff (FDR < 0.05) (Trapnell et al., 2012). The DEGs were identified through the comparison of gene expression among the Yellow, Orange, Brown, and Black stages of pods for each genotype. The calculated P-value was then adjusted through a false discovery rate (FDR)

correction. Genes with adjusted P-values smaller than 0.05 were considered to be significantly differentially expressed.

4.4 Results

4.4.1 Illumina Sequencing and Genome-guided assembly

In order to gain an overall and for initial comparison of peanut transcriptomes in different pod growth stages, the four different genotypes, AABB, AAbb, aaBB and aabb were designed for paired-end (PE) sequencing. After filtering the raw reads, RNA-seq of 16 samples of the four genotypes with three replicates under different growth stages generated a total of 78.57G of 100-bp cleaned reads with an average of 48.78 million read pairs per library (Table 4.1). High-quality reads, with GC percentage varying between 41% to 49%, were obtained, stating superior coverage across the whole transcriptome. After trimming, 84.8% of the raw reads were survived (Table 4.1). The cleaned reads were mapped to the cultivated peanut genome, and the overall mapping rate per library ranged from 62.10% to 77.10%, with an average mapping rate of 66.67% (Table 4.1).

Through the genome-guided assembly, a total of 74,420 genes were assembled for genotype AABB, 74,510 genes were assembled for AAbb, 77,160 genes were assembled for aaBB, and 74,150 genes were assembled for aabb, respectively. There were 66,986 (90.01%), 67,066 (90.01%), 67,159 (87.04%) and 67,058 (90.44%) assembled genes were matched to genes annotated from the cultivated peanut reference genome in genotype AABB, AAbb, aaBB and aabb, respectively (Table 4.2). As a result, 7,434, 7,444, 10,001, and 7,092 novel genes were identified in genotypes of AABB, AAbb, aaBB and aabb, respectively.

4.4.2 Analysis of differentially expressed genes

The DEGs were determined among four pod growth stages of each genotype to study the expression pattern in different stages, as well as among four genotypes of each growth stages to

study the expression pattern in different genotypes. For pod growth stages, there are a totally of 356 genes differentially expressed in the Yellow stage, 1,406 genes in the Orange stage, 732 genes in the Brown stage, and 1,519 genes in the Black stage, respectively (Table 4.2). Of the 369 genes in the Yellow stage, the expression of 157 genes was increased up-regulated, and the level of the other 212 genes were downregulated. For the Orange stage, 609 genes were up-regulated, and 797 genes were downregulated (Table 4.2). For the late pod growth stage, 732 DEGs were identified in Brown stage, including 388 up-regulated genes and 344 down-regulated genes. In addition, a total of 1,519 DEGs were identified in Black stage including 949 up-regulated genes and 570 down-regulated genes. Among the DEGs identified, 218, 795, 362 and 739 DEGs were annotated with the reference genome in Yellow, Orange, Brown and Black stage, respectively (Table 4.2). There are totally 526 genes were differentially expressed in genotype AABB, as well as 923 genes, 2,286 genes, and 758 genes AAbb, aaBB and aabb, respectively (Table 4.2). For genotype AABB, which has a normal level of oleic acid, 231 genes were up-regulated, and 465 genes were downregulated (Table 4.2). Of the 923 genes in genotype AAbb, the expression of 418 genes was up-regulated, and the level of the other 505 genes were downregulated. Besides, 2286 DEGs were identified in genotype aaBB, including 790 up-regulated genes and 1,496 down-regulated genes. Furthermore, a total of 758 DEGs were identified in genotype aabb, which has a high level of oleic acid, including 327 up-regulated genes and 431 down-regulated genes. Among the DEGs identified,401, 534, 1,584, and 569 DEGs were annotated with the reference genome in genotype AABB, AAbb, aaBB and aabb, respectively (Table 4.2).

Pairwise comparison of the DEGs from the four genotypes was performed to investigate which genes failed to respond to the oleic acid metabolic pathway in double-mutant genotypes aabb, single mutant AAbb, and aaBB, as well as the wild genotypes AABB (Fig. 4.1). Likewise, the

pairwise comparison of the DEGs was also conducted among four pod growth stages, including two early mature stages (Yellow and Orange stages) and two late mature stages (Brown and Black stage) (Fig. 4.1). A total of 65 DEGs were shared by pod growth stages (Fig. 4.1).

4.4.3 DEGs located within 3 Mb of significant SNPs associated with O/L ratio

To further study the metabolic regulation pathway of oleic acid in peanut, a comparison analysis was conducted between significant DEGs and SNPs associated with the fatty acid profile by GWAS analysis within 3Mb. Thereby, 31 significant DEGs associated with oleic acid were confirmed by GWAS analysis, including 7 DEGs located in sub-genome A, and 25 DEGs located in sub-genome B. Further sequence location study found out that all DEGs in sub-genome A was identified in chromosome A05, including L5R9I5, 4E075G, 4GH0FM, V10K31, ZL7D6P, and NBP004 while DEGs in sub-genome B dispersed in B03 (6WM8G7, T1VTNT, TZR4FL, WG9HC3), B04 (48P5MU3, 6G1FH3, XAZ91P), B05 (5EZ2U4, J8E4BG, 6J07HI), B09 (09EZ08, Q8L9VW, 73DSWE, N8YXAN, 5F0RSX), and B10 (H8XLFU, ZF9UF1, A0V2WM, RTH5QB, LAL591, 06P6RC, 6ZP1HE, 8GXM4F, FYY1I0). It is noticeable that three DEGs located in Chromosome B04 were found to locate 0.08 Mb from SNP marker AX-176800740, which was identified to be associated with traits IV and O/L ratio in fatty acid profile using GWAS analysis (Table 4.3). Furthermore, this SNP marker was found to locate 2.11 Mb from SSR marker pPGSseq13A10, which was also identified to be associated with O/L ratio (Table 3.9)

4.5 Discussion

Transcriptome profile is an estimable resource for discovering gene expression levels, characterizing new alleles, and developing molecular markers in plants. However, studies on the transcriptome of peanut, especially on sensory and quality attributes analysis, are limited.

Especially for peanut, a non-model organism, the absence of a high-quality reference genome makes the study more challenging.

In this research, a total of 16 assemblies were produced, with four samples from each genotype under different pod growth stages through various assemblers and analytical approaches . reported 74,420, 74,510, 77,160 and 74,150 genes for genotypes AABB, AAbb, aaBB, and aabb with 526, 923, 2,286 and 758 DEGs in each genotype, respectively. The majority of the DEGs were involved in fatty acid metabolic or synthesis process. The result of the Venn diagram and heat map showed gene expression pattern was changed along with the plant growth stage as well as mutation of critical enzymes, ahFAD2A, and ahFAD2B (Fig 4.1 & Fig 4.2).

Furthermore, comparison analysis was conducted between DEGs and SNPs associated with the fatty acid profile, and we obtained 31 significant DEGs associated with oleic acid. Interestingly, three DEGs in Chromosome B04 were found to locate 0.08 Mb from SNP marker AX-176800740 using GWAS analysis, which was also found to locate 2.11 Mb from SSR marker pPGSseq13A10, associated with O/L ratio (Table 3.9 &Table 4.3). Among these three genes, 8P5MU3 was annotated to code a protein of function unknown, while 6G1FH3 and XAZ91P were reported as Peroxidase superfamily protein and ubiquitin carboxyl-terminal hydrolase 27 –like, respectively, and involved in the oxidation-reduction process and process ubiquitination process, respectively. Consistently, it was reported that peanuts with high oleic acid contents were determined to have improved stability against lipid oxidation that could lead to adverse flavors (Talcott et al., 2005), indicating a complicated pathway in fatty acid regulation.

Overall, as the Next-Generation Sequencing technologies possessing more advantages, such as lower-costs and higher-throughput, it's becoming a reality to efficiently achieving the de novo genome assembly, molecular markers development and genome diversity studies, as well as novel

genes identification and gene expression patterns exploration, especially using the RNA-seq strategies (Varshney et al., 2009). In our study, we sequenced and characterized the peanut transcriptomes of different genotypes. The present study demonstrated that different growth stages alter the transcriptome profile in four peanut genotypes with varying oleic acid levels. Thousands of novel genes of cultivated peanuts were identified and annotated. The DEGs involved in fatty acid metabolism and the biosynthesis process were enriched. Besides, the assembly strategies in our study not only were proved to have an efficient and effective potential to assemble the whole RNA sequences with a higher precision and provide a more delicate and more detailed picture of the transcriptome, but also would be deemed to be employed for further differential gene expression analysis, association mapping and QTL verification using various molecular genetic markers, such as SNPs and SSPs in our research, in the cultivated peanut. This study provided not only insights into putative peanut regulation in different genotypes under different growth stages, but also a piece of valuable transcriptomics information to smooth functional genomics studies for future peanut breeding.

4.6 References

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Bertioli, D. J., Jenkins, J., Clevenger, J., Dudchenko, O., Gao, D., Seijo, G., … Pandey, M. K. (2019). The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. Nature Genetics, 51(5), 877.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Chen, X., Lu, Q., Liu, H., Zhang, J., Hong, Y., Lan, H., … Liang, X. (2019). Sequencing of Cultivated Peanut, *Arachis hypogaea*, Yields Insights into Genome Evolution and Oil Improvement. Molecular Plant, 12(7), 920–934. https://doi.org/10.1016/j.molp.2019.03.005

Chen, X., Zhu, W., Azam, S., Li, H., Zhu, F., Li, H., … Liang, X. (2013). Deep sequencing analysis of the transcriptomes of peanut aerial and subterranean young pods identifies candidate genes related to early embryo abortion. Plant Biotechnology Journal, 11(1), 115–127. https://doi.org/10.1111/pbi.12018

Chi, X., Yang, Q., Chen, X., Wang, J., Pan, L., Chen, M., … Yu, S. (2011). Identification and characterization of microRNAs from peanut (*Arachis hypogaea* l.) by high-throughput sequencing. PLoS ONE, 6(11). https://doi.org/10.1371/journal.pone.0027530

Chopra, R., Burow, G., Farmer, A., Mudge, J., Simpson, C. E., & Burow, M. D. (2014). Comparisons of de novo transcriptome assemblers in diploid and polyploid species using peanut (*Arachis* spp.) RNA-Seq data. PLoS ONE, 9(12), 1–16. https://doi.org/10.1371/journal.pone.0115055

Deng, X., Li, Z., & Zhang, W. (2012). Transcriptome sequencing of Salmonella enterica serovar Enteritidis under desiccation and starvation stress in peanut oil. Food Microbiology, 30(1), 311–315. https://doi.org/10.1016/j.fm.2011.11.001

Guimaraes, P. M., Guimaraes, L. A., Morgante, C. V., Silva, O. B., Araujo, A. C. G., Martins, A. C. Q., … Brasileiro, A. C. M. (2015). Root transcriptome analysis of wild peanut reveals candidate genes for nematode resistance. PLoS ONE, 10(10), 1–22. https://doi.org/10.1371/journal.pone.0140937

Kellner, F., Kim, J., Clavijo, B. J., Hamilton, J. P., Childs, K. L., Vaillancourt, B., … O'Connor, S. E. (2015). Genome-guided investigation of plant natural product biosynthesis. Plant Journal, 82(4), 680–692. https://doi.org/10.1111/tpj.12827

Lang, W., Cai, J., Chen, F., Zhu, J., Hu, X., & Zhong, J. (2019). RNA-Seq analyses demonstrate EVI-1–induced morbid hematopoiesis and developmental abnormality in zebrafish were related with MAPK pathway. Hematological Oncology, 37(3), 326–329. https://doi.org/10.1002/hon.2609

Li, B., Tsoi, L. C., Swindell, W. R., Gudjonsson, J. E., Tejasvi, T., Johnston, A., … Elder, J. T. (2014). Transcriptome analysis of psoriasis in a large case-control sample: RNA-seq provides insights into disease mechanisms. Journal of Investigative Dermatology, 134(7), 1828–1838. https://doi.org/10.1038/jid.2014.28

Lu, X., Li, J., Yang, J., Liu, X., & Ma, J. (2014). De novo transcriptome of the desert beetle Microdera punctipennis (Coleoptera: Tenebrionidae) using illumina RNA-seq technology. Molecular Biology Reports, 41(11), 7293–7303. https://doi.org/10.1007/s11033-014-3615-6

Janila, P., Pandey, M. K., Shasidhar, Y., Variath, M. T., Sriswathi, M., Khera, P., … Varshney, R. K. (2016). Molecular breeding for introgression of fatty acid desaturase mutant alleles (ahFAD2A and ahFAD2B) enhances oil quality in high and low oil containing peanut genotypes. Plant Science, 242, 203–213. https://doi.org/10.1016/j.plantsci.2015.08.013

Pasakarnis, T. S., Gorski, C. A., O'Loughlin, E., Parkin, G. F., & Scherer, M. M. (2006). Abiotic reduction of chlorinated ethenes in the presence of anaerobic bacteria. ACS, Division of Environmental Chemistry - Preprints of Extended Abstracts, 46(1), 957–961.

Ruan, J., Guo, F., Wang, Y., Li, X., Wan, S., Shan, L., & Peng, Z. (2018). Transcriptome analysis of alternative splicing in peanut (*Arachis hypogaea* L.). BMC Plant Biology, 18(1), 1–11. https://doi.org/10.1186/s12870-018-1339-9

Shalek, A. K., Satija, R., Shuga, J., Trombetta, J. J., Gennert, D., Lu, D., … Regev, A. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature, 510(7505), 363–369. https://doi.org/10.1038/nature13437

Strauß, T., Van Poecke, R. M. P., Strauß, A., Römer, P., Minsavage, G. V., Singh, S., … Lahaye, T. (2012). RNA-seq pinpoints a Xanthomonas TAL-effector activated resistance gene in a large-crop genome. Proceedings of the National Academy of Sciences of the United States of America, 109(47), 19480–19485. https://doi.org/10.1073/pnas.1212415109

Talcott, S. T., Passeretti, S., Duncan, C. E., & Gorbet, D. W. (2005). Polyphenolic content and sensory properties of normal and high oleic acid peanuts. Food Chemistry, 90(3), 379–388. https://doi.org/10.1016/j.foodchem.2004.04.011

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., … Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature Protocols, 7(3), 562.

Vassiliou, E. K., Gonzalez, A., Garcia, C., Tadros, J. H., Chakraborty, G., & Toney, J. H. (2009). Oleic acid and peanut oil high in oleic acid reverse the inhibitory effect of insulin production of the inflammatory cytokine TNF- both in vitro and in vivo systems. Lipids in Health and Disease, 8, 1–10. https://doi.org/10.1186/1476-511X-8-25

Yin, D., Liu, H., Zhang, X., & Cui, D. (2011). A protocol for extraction of high-quality RNA and DNA from peanut plant tissues. Molecular Biotechnology, 49(2), 187–191. https://doi.org/10.1007/s12033-011-9391-9

Zhao, C. Z., Xia, H., Frazier, T. P., Yao, Y. Y., Bi, Y. P., Li, A. Q., … Wang, X. J. (2010). Deep sequencing identifies novel and conserved microRNAs in peanuts (*Arachis hypogaea* L.). BMC Plant Biology, 10, 1–12. https://doi.org/10.1186/1471-2229-10-3

Zhao, Q., Wu, J., Cai, G., Yang, Q., Shahid, M., Fan, C., … Zhou, Y. (2019). A novel quantitative trait locus on chromosome A9 controlling oleic acid content in Brassica napus. Plant Biotechnology Journal, 17(12), 2313–2324. https://doi.org/10.1111/pbi.13142

Zhao, S., Chen, A., Chen, C., Li, C., Xia, R., & Wang, X. (2019). Transcriptomic analysis reveals the possible roles of sugar metabolism and export for positive mycorrhizal growth responses in soybean. Physiologia Plantarum, 166(3), 712–728. https://doi.org/10.1111/ppl.12847

Table 4.1 Summary of library, trimming and alignment of reads to A. *hypogea* genome in each library.

Growh Stage	Genotype	Input read pairs	Both surviving	Forward only surviving	Reverse only surviving	Dropped	Overall read mapping rate (%)	Overall alignment rate (%)
Black	**aabb**	45105035	39328733 (87.19%)	3033856 (6.73%)	1003348 (2.22%)	1739098 (3.86%)	97.5	63.2
Brown	**aabb**	43434238	34507782 (79.45%)	4332480 (9.97%)	4079798 (9.39%)	514178 (1.18%)	97.5	62.1
Orange	**aabb**	51764377	45120235 (87.16%)	3388616 (6.55%)	1274307 (2.46%)	1981219 (3.83%)	97.8	65
Yellow	**aabb**	53164650	45047730 (84.73%)	4680841 (8.80%)	1040556 (1.96%)	2395523 (4.51%)	97.8	66.6
Black	**aaBB**	45004462	39179183 (87.06%)	3008286 (6.68%)	1041719 (2.31%)	1775274 (3.94%)	98	65.6
Brown	**aaBB**	38052751	31561076 (82.94%)	3660616 (9.62%)	768365 (2.02%)	2062694 (5.42%)	96.1	64.7
Orange	**aaBB**	45471555	39346736 (86.53%)	3321344 (7.30%)	1000206 (2.20%)	1803269 (3.97%)	97.5	63.8
Yellow	**aaBB**	45240146	37276439 (82.40%)	3702368 (8.18%)	1664996 (3.68%)	2596343 (5.74%)	97.8	77.1
Brown	**AAbb**	41830491	34750271 (83.07%)	3217459 (7.69%)	1498919 (3.58%)	2363842 (5.65%)	98	69.4
Orange	**AAbb**	64137923	54527200 (85.02%)	4851027 (7.56%)	1590888 (2.48%)	3168808 (4.94%)	97.2	66.7
Yellow	**AAbb**	44620537	38210547 (85.63%)	3298932 (7.39%)	968312 (2.17%)	2142746 (4.80%)	97.9	71.8
Black	**AABB**	51344509	43214611 (84.17%)	4524714 (8.81%)	1072371 (2.09%)	2532813 (4.93%)	95.9	64.3
Brown	**AABB**	55170129	47283787 (85.71%)	4086040 (7.41%)	1274297 (2.31%)	2526005 (4.58%)	97.6	62.9
Orange	**AABB**	62960648	53811304 (85.47%)	4769960 (7.58%)	1617414 (2.57%)	2761970 (4.39%)	96	65.9
Yellow	**AABB**	44386707	37950130 (85.5%)	3321070 (7.48%)	1133103 (2.55%)	1982404 (4.47%)	97.4	69.9

Table 4.2 Summary of library and alignment of reads to *A. hypogea* genome in each library.

Genotype	Total Genes	Unannotated	Annotated	Total DEGs	Unannotated	Annotated
aabb	74,150	7,092	67,058	758	189	569
aaBB	77,160	10,001	67,159	2,286	702	1,584
AAbb	74,510	7,444	67,066	923	389	534
AABB	74,420	7,434	66,986	526	125	401

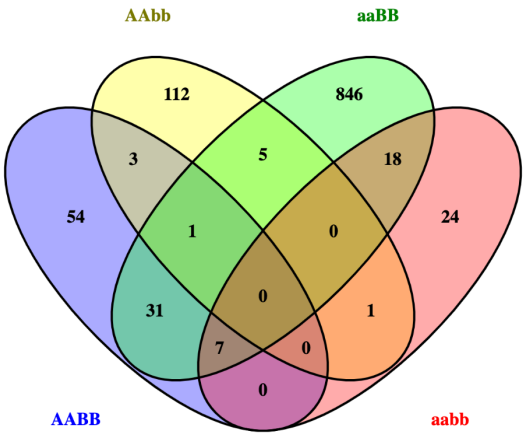
Growth Stage	Total Genes	Unannotated	Annotated	Total DEGs	Unannotated	Annotated
Yellow	74,966	7,917	67,049	369	151	218
Orange	75,346	8,274	67,072	1,406	611	795
Brown	75,578	8,528	67,050	732	370	362
Black	75,226	8,165	67,061	1,519	780	739

Table 4.3 Summary of DEGs identified within 3 Mb window of significant SNPs by GWAS analysis.

SNP ID	Chr	Locus	DEGs	Position (bp)		Note
AX-176811155	A5	95722033	L5R9I5	94691319	94693338	pyridoxal phosphate binding
AX-176811155	A5	95722033	4E075G	94754736	94760831	Regulator of Vps4 activity in the MVB pathway protein
AX-176811155	A5	95722033	4GH0FM	94840510	94844874	Unknown protein
AX-176798965	A5	35912339	V10K31	35104929	35109540	diacylglycerol O-acyltransferase activity
AX-176798965	A5	35912339	ZL7D6P	35301085	35303209	Aminotransferase like plant mobile domain
AX-176798965	A5	35912339	NBP004	35991927	35992857	Unknown protein
AX-176798965	A5	35912339	AE9D0W	36016579	36019770	protein phosphorylation
AX-176816801	B3	9686512	6WM8G7	9907886	9909872	oxidation reduction process
AX-176816801	B3	9686512	T1VTNT	10402032	10403890	hypothetical protein
AX-176816801	B3	9686512	TZR4FL	10666107	10668098	transporter activity
AX-147243963	B3	18642606	WG9HC3	17850190	17855872	transmembrane transport
AX-176800740	B4	118083731	8P5MU3	117969930	117971196	Protein of unknown function DUF506 plant
AX-176800740	B4	118083731	6G1FH3	117986388	117999512	oxidation-reduction process
AX-176800740	B4	118083731	XAZ91P	119241737	119244212	ubiquitin carboxyl terminal hydrolase 27 -like
AX-176800740	B4	118083731	XAZ91P	119241740	119244212	ubiquitin carboxyl terminal hydrolase 27 -like
AX-176819943	B5	134187510	5EZ2U4	133132918	133134113	transmembrane protein putative
AX-176819943	B5	134187510	J8E4BG	133506059	133511819	potassium ion transmembrane transport
AX-176819943	B5	134187510	6J07HI	134596895	134602246	integral component of membrane
AX-177643501	B9	10735224	09EZ08	9736781	9741180	electron carrier activity

AX-177643501	B9	10735224	Q8L9VW	10179439	10184128	uncharacterized protein
AX-177643501	B9	10735224	73DSWE	10808033	10814915	integral component of membrane
AX-177643084	B9	49917116	N8YXAN	49573891	49582783	Breast carcinoma amplified sequence 3
AX-177643084	B9	49917116	5F0RSX	49889649	49893023	hypothetical protein
AX-176791500	B10	4946993	H8XLFU	4344812	4345545	Dormancy/auxin associated family protein
AX-176791500	B10	4946993	ZF9UF1	5030711	5031441	signal transduction
AX-176791500	B10	4946993	A0V2WM	5321452	5323079	Protein of unknown function
AX-176791500	B10	4946993	RTH5QB	5713680	5719071	integral component of membrane
AX-176804144	B10	115730891	LAL591	14859083	14864512	unknown protein
AX-176804144	B10	115730891	06P6RC	14972628	14975225	mitotic spindle assembly checkpoint
AX-177638452	B10	117191238	6ZP1HE	117679565	117682849	Ulp1 protease family carboxy
AX-177638677	B10	120214136	8GXM4F	120016130	120018502	MTD1 D2 Tax Medicago truncatula Rep
AX-177639314	B10	124890936	FYY1I0	125034987	125040016	integral component of membrane

A



B

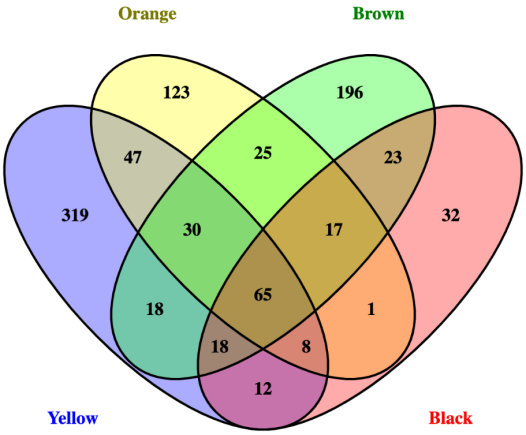
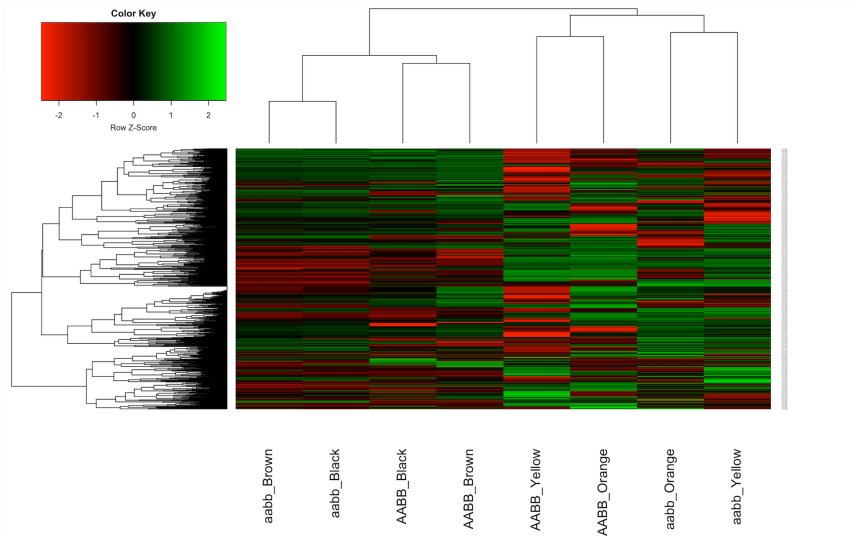


Figure 4.1 Comparison of the annotated DEGs among A) different genotypes including AABB, AAbb, aaBB, and aabb; and B) different growth stages, including Yellow, Orange, Brown, and Black stage.

A



B

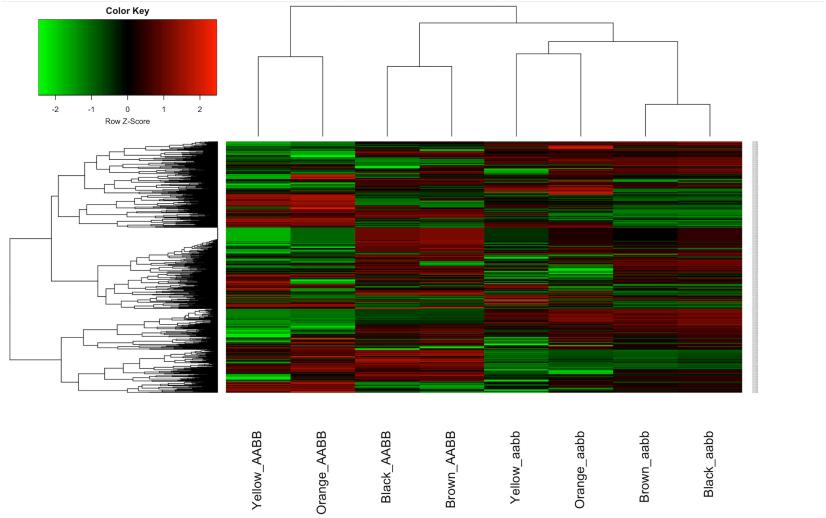


Figure 4.2 Expression profiles of the differentially expressed genes by A) All four genotypes under different growth stages; and B) All four growth stages in different genotypes. Log10 transformed FPKM values were used. "green" color indicates low expression level and "red" color indicates high expression level.

Chapter Five

Research Summary and Future Perspectives

5.1 Research Summary

The research projects mainly focused on the investigation of the identification of molecular markers using for Marker-assisted Selection, including SNPs and SSRs. In our study, a total of 276 significantly QTLs associated with four main sensory and quality attribute was identified, and consequently, 136 candidate genes were screened functionally associated with the corresponding traits. This is the first GWAS analysis in US Mini-core collections that identified sensory and quality attributes related to QTLs and the novel candidate genes which have never been found to affect flavor and nutrition in peanut. Fine genetic mapping of the QTL associated with sensory and quality attributes will allow the application of marker-assisted selection and understanding of underlying molecular mechanisms. Moreover, further verification and study were conducted to demonstrate the candidate genes having functions in the biological metabolic process in cultivated peanut, including association mapping using SSRs and RNA-Seq analysis. During association mapping using SSRs, we found out that there were 37 significantly associated SSRs identified from association mapping, confirmed by GWAS analysis, including three SSRs associated with flavors profile, 10 SSRs associated with sugars profile, 11 SSRs associated with fatty acids profile, and 13 SSRs associated with tocopherols profile. Also, 31 DEGs were identified within 3 Mb window of significant SNPs associated with corresponding traits identified through GWAS analysis. Besides, during RNA-seq analysis, both unique and well-known sensory and quality attributes related genes were identified in this study.

5.2 Future Perspectives

The results achieved from these three studies provide distinguished genetic materials and molecular markers, including SNPs and SSRs for peanut breeding in sensory and quality attributes, as well as understand the sensory and quality attributes regulatory mechanism of peanut. It would definitely promote future efforts attempting to breed flavor-desirable and nutrient-rich peanut genotypes. However, further efforts are still needed to investigate the molecular signaling and regulatory mechanisms of peanut due to the complex genetic background. More verification and study such as Quantitative Real-time PCR or genetic knockout lines analysis should be performed to characterize their possible functional roles in sensory and quality attributes related to the biological metabolic process in cultivated peanut. After confirmation of the possible functions of the flavor related genes, targeted gene modification might be carried out to improve and accelerate peanut breeding. Besides, with the verified molecular markers, Marker-assisted Selection could also be implemented to obtain flavor-desirable and nutrient-rich peanut. Overall, these findings would provide encouraging insight into the complicated genetic architecture and molecular regulation mechanisms of sensory and quality attributes, and promote the Marker-assisted Selection in peanut breeding with seed flavor and nutrition quality.