CHARACTERIZATION OF A PEANUT MULTIPARENTAL ADVANCED GENERATION INTERCROSS (PEANUTMAGIC) POPULATION AND CONSTRUCTION OF A POPULATION-SPECIFIC PANGENOME FOR GENOMIC STUDIES AND TRAIT MAPPING

by

ETHAN ALDEN THOMPSON

(Under the Direction of Albert K. Culbreath)

ABSTRACT

Peanut (*Arachis hypogaea* L.) is a globally significant crop valued for its oil and protein content, making cultivars with high seed quality and disease resistance desirable. Genetic mapping of these traits relies on both a suitable population and accurate genotyping, however, these efforts have been hindered by the narrow genetic diversity of cultivated peanut. The objective of this dissertation is to evaluate the utility and power of an eight-way peanut multiparent advanced generation intercross (PeanutMAGIC) population in genomic studies and to compare single-reference and population-specific pangenome genotyping for PeanutMAGIC. A random subset of 310 recombinant inbred lines (RILs), termed the MAGIC Core, was selected out of the whole population of 3,187 (RILs) for phenotyping and genotyping. The MAGIC Core was phenotyped for pod and seed traits for initial population characterization and disease resistance for additional application of the population, including oleic acid, *Tomato Spotted Wilt Virus* (TSWV), leaf spots, and root-knot nematodes. The genomes of the eight founders of PeanutMAGIC were sequenced to construct a population-specific graph-based pangenome. This pangenome enabled a

comprehensive library of over 2.7 million variants present in the population to empower genomic studies and high-resolution trait mapping. The MAGIC Core was subject to low-coverage whole-genome sequencing for single-reference and population-specific pangenome reference genotyping. The characterization study of the MAGIC Core, using pod and seed traits with single-reference genotyping calling, demonstrated that PeanutMAGIC was a well-balanced population with increased recombination that is suitable for genetic and genomic studies and breeding. Three long-standing-sought-after traits have been mapped using pangenome-based markers in genome wide association studies (GWAS), that were not found in single-reference genotyping: (1) revealed a novel third fatty acid desaturase gene (*AhFAD*2C) required for high oleic acid peanut and resolved the long-standing mystery of mid-oleic acid phenotypes; (2) identified a copy number variant (CNV) exclusive to the TSWV resistant founder 'NC94022' that contains tandem duplications of a novel glutamate receptor; (3) narrowed the root-knot nematode resistance locus to a 400kb region. These findings underscore the power of population-specific pangenomics in enhancing marker resolution, uncovering structural variants, and empowering genomic studies and trait associations.

INDEX WORDS: Peanut, MAGIC, Multiparental, Genomics, Pangenomics, Fatty Acid Desaturase, Tomato Spotted Wilt Virus, Leaf Spot, Root Knot Nematode

CHARACTERIZATION OF A PEANUT MULTIPARENTAL ADVANCED GENERATION INTERCROSS (PEANUTMAGIC) POPULATION AND CONSTRUCTION OF A POPULATION-SPECIFIC PANGENOME FOR GENOMIC STUDIES AND TRAIT MAPPING

by

ETHAN ALDEN THOMPSON

B.S. North Carolina State University, 2021

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHIILOSOPHY

ATHENS, GEORGIA

2025

CHARACTERIZATION OF A PEANUT MULTIPARENTAL ADVANCED GENERATION INTERCROSS (PEANUTMAGIC) POPULATION AND CONSTRUCTION OF A POPULATION-SPECIFIC PANGENOME FOR GENOMIC STUDIES AND TRAIT MAPPING

by

ETHAN ALDEN THOMPSON

Major Professor: Albert K. Culbreath

Committee: Baozhu Guo
Josh P. Clevenger
Peggy Ozias-Akins
David Bertioli
Soraya Bertioli

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
August 2025

DEDICATION

To my lovely bride Makenzie. To my pretty girl Veronica, you will be missed.

ACKNOWLEDGEMENTS

Thank you, heavenly father for the opportunity to go on this journey and for it to serve you, how you see fit. I have matured and developed greatly during this time and am better off because of it. I will use this experience to continue to serve your will.

Thank you to my lovely bride who has been here throughout this journey as my support at home. Your love and value are unparalleled.

Thank you, Dr. Guo, for facilitating an environment to develop not only my scientific career through resources and connections but my character through your wisdom and life stories. For there are many scientists that can serve as mentors, but few will care for you like a father.

Thank you, Dr. Clevenger, for your generosity and expertise to drive innovation not only for my personal development, but for the peanut research community.

Thank you, Dr. Culbreath, for serving as my major professor, guiding my understanding of peanut pathogens, and helping me traverse the departmental requirements.

Thank you to David Bertioli, Soraya Bertioli and Peggy Ozias-Akins for presenting resources and challenges to develop my scientific mind, writing, and knowledge.

Thank you to the unsung heroes that have helped me in the lab, field, paperwork, and in the cluster. From Dr. Guo's Lab: Billy, Wu, Hui, Caleb, Steven, Makayla, Taylor, Noah, Ballen, Cody, Sunil, and Wyatt; Dr. Peggy's lab: Stephanie, Sydney, and Akshaya; Dr. Clevenger's lab: Walid, Kendell, Zack, Paige, TJ, and Abby; glue of the department: Christy and Tara. There have been many hurdles that I have needed help with, and I could not complete without you.

GO DAWGS!

TABLE OF CONTENT

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Genetic Mapping Approaches in Peanut

Cultivated peanut (*Arachis hypogaea* L.) is an important legume crop grown worldwide. However, cultivated peanut has a narrow genetic base with limited genetic diversity because of its origin from a recent hybridization and genome duplication of two diploid species, *Arachis duranensis* (A-genome) and *Arachis ipaënsis* (B-genome) (Bertioli et al., 2016, 2019). Cultivated peanut genomic architecture is allotetraploid (2n = 4x = 40) with two sets of chromosome (Chr.) pairs (AABB; AA = Chr. 1–10, BB = Chr. 11–20), where wild diploid relatives (2n = 2x = 20) have mostly AA- or BB-type genomes (Bertioli et al., 2016, 2019), restricting the movement of wild alleles into the cultivated peanut germplasm. As early as 1936, Husted (Husted, 1936) hypothesized that cultivated peanut developed from a single hybridization of two diploid progenitors, which was supported by recent genomic technologies (Bertioli et al., 2019; Seijo et al., 2004). The low genetic diversity of cultivated peanut has limited genetic and genomic studies in the peanut community (Guo et al., 2012, 2016), despite the continual efforts of peanut researchers, like breeders (Brown et al., 2021; Isleib et al., 2001), to broad genetic diversity by bringing new germplasm into breeding programs. These efforts have been driven by the need for disease resistance introduced through germplasm and interspecific hybridization with diploid relatives (Isleib et al., 2001).

In the last decade, the worldwide peanut community has witnessed the progress made in genomic research through the availability of high-quality reference genomes for both diploid and

tetraploid peanuts (Bertioli et al., 2016, 2019; Chen et al., 2016, 2019; Zhuang et al., 2019). As the cost of sequencing decreases and accuracy in marker selection increases, particularly with polyploids like peanuts (J. Clevenger et al., 2015; J. P. Clevenger & Ozias-Akins, 2015), marker saturation is no longer a limiting factor in genetic mapping of traits of interest (Agarwal et al., 2019; Korani et al., 2021; Stange et al., 2013). As a result, there is a growing need for populations with greater genetic diversity and recombination density. Recently, the availability of genomic tools designed for repetitive understudied genomes (Korani et al., 2021) has facilitated the dissection of complex populations, making the construction and observation of these populations accessible.

Biparental populations have been the primary focus of plant genetics. However, they are limited by their construction from a single cross involving only two genetic donors (Arrones et al., 2020; Scott et al., 2020). Most peanut genetic mapping studies have used biparental recombinant inbred line (RIL) populations, such as the "S" and the "T" RIL populations (Agarwal et al., 2018, 2019; Khera et al., 2016; Pandey et al., 2017; Qin et al., 2012). These studies also used simple sequence repeat (SSR) markers (Hong et al., 2010; M. Luo et al., 2005), resulting in lower resolution in quantitative trait locus (QTL) prediction (Khera et al., 2016; Pandey et al., 2017). Recent studies by Agarwal et al. (Agarwal et al., 2018, 2019), using the same "S" and "T" RIL populations, demonstrated the power of genotyping by whole genome resequencing (WGRS), which increased the maker density and the ability to track recombination events compared to SSR marker approaches.

Multiparent populations are the converging point between biparental populations and diversity panels. Biparental populations have the advantage of being easy to synthesize and having known pedigrees. However, biparental populations suffer from a lack of diversity and

limited recombination (Arrones et al., 2020; Scott et al., 2020). Diversity panels are the opposite, where there is high genetic diversity and historical accumulation of recombination events in the panel, but they are limited by unknown pedigrees and parental information (Arrones et al., 2020; Dell'Acqua et al., 2015; Scott et al., 2020). Multiparent populations bridge the gap, allowing for the combination of several diverse lines into one population, thus, increasing recombination and diversity while preserving the pedigree information of the population (Dell'Acqua et al., 2015; Scott et al., 2020). There are several multiparental population structures utilized in plant breeding such as nested association mapping (NAM) and multiparent advanced generation intercross (MAGIC) (Arrones et al., 2020; Scott et al., 2020). NAM is a breeding design where a common parent is crossed to an array of other unique parental genotypes. There are limitations to NAM population structures, including a focus on dissecting the common parent's genotype through recombination with a single other founder in each panel. While peanut NAM populations have been constructed (Holbrook et al., 2013), the individual NAM families have been primarily used in a biparental population fashion (Chavarro et al., 2019; Chu et al., 2019; Z. Luo et al., 2020) because of the logistical challenges of working with large number of biparental populations. Gangurde et al. (Gangurde et al., 2020, 2024) reported the power of the NAM approach in mapping pod and seed traits and late leaf spot resistance using a subset of these peanut NAM populations.

In contrast to dissection of a primary genotype in NAM , MAGIC population designs facilitate a segregating population with equal genetic contribution from all founders, thereby enhancing diversity and recombination through multiple crossing generations and distinct haplotypes combinations, resulting in RILs with known pedigrees that improves the capability to dissect complex genetic traits (Arrones et al., 2020; Dell'Acqua et al., 2015; Scott et al., 2020;

Thompson et al., 2024, 2025). During the intercrossing generations, the meiotic recombination events are the essential process that rearranges alleles and produces new combinations of genes for breeding lines. This breeding design supports plant breeding by broadening the genetic base of the population and pyramiding key agronomic traits into lines that can be used in cultivar development. The PeanutMAGIC population exemplifies the advantages of a multiparental breeding approach in peanut, providing increased genetic variation and recombination for genomic studies (Thompson et al., 2024). The increased genetic diversity and recombination in MAGIC populations enable finer genetic mapping of quality traits such as oleic acid content to resolve the mystery that has eluded the peanut community. Additionally, these qualities allow for the dissection of more complex phenotypes such as disease resistance that has been a main objective in peanut research (Agarwal et al., 2018, 2019; Chu et al., 2011, 2016, 2019; J. Clevenger et al., 2017; Culbreath et al., 2005; Gangurde et al., 2024; Khera et al., 2016; Z. Luo et al., 2020; Pandey et al., 2017; Qin et al., 2012; Simpson, 1991; Tallury et al., 2014; Wang et al., 2016; Wu et al., 2025).

Paradigm Shift to Population Specific Pangenomes

The community reference genome of *A. hypogaea* cv. 'Tifrunner' (TR) was a joint effort of the International Peanut Genome Sequencing Consortium (Bertioli et al., 2019). The publication of a reference genome signified a major milestone in the peanut community, due to the allotetraploid (AABB-type genome; 2n = 4x = 40 chromosomes) nature of this species and limited subgenome divergence, stemming from its recent hybridization event from its related progenitors (*A. duranensis*, AA; *A. ipaensis*, BB) (Bertioli et al., 2016, 2019). These factors have made genome sequencing and assembly challenging for tetraploid peanut (Bertioli et al., 2019).

The reference genome has been pivotal for divulging the intricacies and origin of cultivated peanut for crop improvement; however, it does not reveal all the genetic variance within cultivated peanut or a particular mapping population. In whole genome sequencing, sequence reads are collected for an individual and mapped to a reference for genotype calling. When lines have a consistent difference from the reference in a particular location, the difference is interpreted as a variant and is called for those lines. However, in repetitive genomes such as polyploids discernment of these sequences become difficult. This affords the opportunity for reads that are from one location, but similar to a different location to be misaligned and interpreted as a variant when the first region is not represented in the reference. Additionally, these alternative alleles that the reference does not possess must be interpreted from the data due to the lack of a database of other segregating alleles.

Conversely, a species level pangenome offers a library of nearly all variation present throughout the species, more variants than within a specific population. The founders of the PeanutMAGIC population do not fully represent the global diversity of peanut. Therefore, the use of a species level pangenome might include variations that are not possible in the population, potentially introducing error and leading to improper genotyping for RILs. To generate a reference that is specific to a mapping population, a population specific pangenome can be used. A population-specific pangenome facilitates a comprehensive library of segregating alleles within a population that can be referenced to for genotyping to reduce reference bias (Danilevicz et al., 2020; L. Guo et al., 2025; Hou et al., 2024; Liu et al., 2024; Thompson et al., 2025; Vaughn et al., 2022) . This approach generates founder-specific markers that facilitate the ability to track recombination points throughout the breeding scheme precisely and provide insights into deciphering phenotypic consequences. The increase in quality and quantity of markers allows for

a more complete understanding of recombination frequencies within a population to empower association studies and downstream analysis.

To build these population-specific graph-based pangenomes, Minigraph-Cactus can be utilized (Armstrong et al., 2020; Garrison et al., 2018; Hickey et al., 2020, 2023; Liao et al., 2023; Sirén et al., 2021). Minigraph-Cactus is a graph-based pangenome builder that directly builds the pangenome from whole-genome alignments allowing for the graph to represent variation at different scales (Hickey et al., 2023). First the pipeline constructs an initial structural variant (SV) graph using minigraph with variants >50bp. Next each assembly and the reference are mapped back to the SV graph and filtered for spurious alignments, minimizing orthologous sequences aligning to inconsistent locations on different versions of the graph. Since minigraph does not introduce interchromosomal events, the nodes must be split into respective chromosomes. This is done using information retained from nodes being connected to the reference assembly from alignment to the SV graph. If a contig is mapped to nodes from multiple chromosomes then it is assigned to the chromosome which most bases align, if the threshold of the alignment to the node is not met then the contig is removed. Once the chromosomes are split, then graph construction proceeds on each chromosome independently, allowing for parallelism of the graph construction. Base alignment is performed through Cactus that ‘pinches’ the exactly matching bases in pairwise alignments to form an initial sequence graph. The initial sequence graph is then transformed into a Cactus graph that is composed of chains of alignment with the sequence graph and filtered for candidate spurious or incomplete alignments corresponding to chains visited by numerous sequences. The last step reassigns node names so they are unique across the graph and collapses redundant sequences. Additionally, the nodes are replaced with their reverse complement to ensure the reference path is always in a

forward orientation, if necessary. This step also identifies and clips highly repetitive sequences such as centromeres and telomeres because of the lack of unique subsequences for alignment seeds. These steps generate a graph based pangenome indexed using the reference coordinates, suitable for studying genomic variations across the input assemblies. By using the founder assemblies of a population, this generates a pangenome specifically for the population that describes the variations between founder lines.

The tools that are available for mapping and variant calling pangenomes have been designed for diploid genomes, making their effectiveness in allotetraploid peanut unclear. Thus, a custom pipeline is needed to accurately identify callable markers specifically for a complex, duplicated genome such as cultivated peanut. An expansion of Khufu was created to use graph based pangenomes and to identify suitable markers, which in theory parallels the current single reference-based marker calling pipeline in Khufu (Korani et al., 2021; Roy et al., 2023). KhufuPAN requires a graphical fragment assembly (GFA) file. The file should include the genomes forming the pangenome and a reference genome, which is used to assign the coordinates of SNPs and structural variants. The first step in the environment is KhufuPAN bootstrapping. The GFA is deconstructed to produce a parental VCF file. Then a series of filters is applied to remove likely erroneous alleles, low-quality variants, or those having polymorphism only with the coordinate reference genome to generate a Filtered-Variant set. All files are packed into a single folder, which can be reused for multiple purposes. KhufuVAR is a KhufuPAN sub-tool used to call and filter GFA variants for markers. Raw FASTQ files, short or long-reads, are applied to Khufu-core (https://www.hudsonalpha.org/khufudata/), and then mapped to the GFA file using vg giraffe (Sirén et al., 2021). Calls under the minimum depth cutoff are masked, and variants overlapping with the Filtered-Variant set are extracted. Variants are segregated

differently based on the population structure. Therefore, another series of filters are applied, i.e., minor allele frequency (>0.01), polymorphism, and the percentage of missing data (>75% missing). To efficiently utilize computational resources, KhufuVAR splits the process into two parts. Within the first part, samples run in batches, and every batch is aligned with the Filtered-Variant set and calculates measurements for the population filters. The second part combines the batches and runs the population filters. The final output calls are exported in Khufu panmap format, which facilitates pangenome-based markers for genomic analysis and associations.

The Mid-Oleic Acid Mystery

Arachis hypogaea L. is cultivated globally as a sustainable and affordable oil and protein source, yielding 54 million tons annually (http://www.fao.org/faostat, 2020). Peanut seeds contain considerable amounts of oil (~50%, dry weight) and protein (~25%, dry weight) that are bioavailable and beneficial for human health (Alper & Mattes, 2003; Jones et al., 2014; Kris-Etherton et al., 2008). Oleic and linoleic fatty acids are two heart-healthy, major components of the total oil content in peanut seeds. Therefore, peanut seed oil composition is an important quality trait for confectionery and wholesale markets, leading to a focus on oil composition in breeding programs and functional studies (Chu et al., 2011; Davis et al., 2016). High-oleic (>75%) peanut cultivars are desirable for their increased shelf life and improved palatability in whole or partial peanut products, whereas low (<55%) to mid-oleic (55-75%) cultivars, with high-palmitic acid content, are valued for peanut butter-based products.

In cultivated peanut, the enzyme fatty acid desaturase 2 (AhFAD2) is responsible for converting oleic acid into linoleic acid, with the functional genes *AhFAD2A* and *AhFAD2B* (Jung et al., 2000; López et al., 2000). Loss-of-function mutations in *AhFAD2A* and *AhFAD2B* have

been identified in breeding populations, with selectable markers available for marker-assisted selection of high-oleic phenotypes (Chu et al., 2007, 2009, 2011). The *AhFAD2A* and *AhFAD2B* markers have been successfully utilized in the development of a high-oleic and root-knot nematode-resistant variety, 'TifNV High O/L', suggesting that tracing two genes are sufficient to predict and select for high-oleic varieties (Chu et al., 2011). However, mid-oleic peanut lines have been observed, but their inheritance could not be explained solely by the previously established two-gene model (Branch et al., 2022), leading to the hypothesis that a third recessive gene, along with AhFAD2A and AhFAD2B, is needed to maintain the high-oleic fatty acid trait in peanut (Branch et al., 2022). Interestingly, Pandey et al. (Pandey et al., 2014) reported that AhFAD2B had a higher phenotypic effect on oleic acid content than AhFAD2A using two biparental populations, and the normal distribution of oleic fatty acid in the mapping populations did not fit the two-gene model (Pandey et al., 2014). The mystery of mid-oleic has perplexed the peanut community, particularly concerning the confectionery industry's need for consistent quality in products. We hypothesized that PeanutMAGIC and a population specific pangenome have the capacity to resolve this mystery to improve our understanding of this important quality trait.

The Elusive TSWV Resistance Loci

Because of the economic importance of peanuts as a whole food product and ingredient for many commodities such as peanut oil, peanut butter and candies, there is a constant interest in improving the peanut germplasm in terms of seed quality, resistance to stresses and commercial productivity. *Tomato spotted wilt virus* (TSWV) threatens the existence of peanut cultivation, which is primarily managed through host-resistance with a required integrated pest

management system. Therefore, TSWV resistance has been a long-sought-after trait for efficient genetic gain, facilitating improved resources for TSWV management and production (Agarwal et al., 2019; Khera et al., 2016; Pandey et al., 2017; Qin et al., 2012; Tseng et al., 2016; Wu et al., 2025). ‘NC94022’ has been identified to possess high levels of field resistance that could be utilized in peanut production (Culbreath et al., 2005). Previous findings have identified both the middle and beginning of Chr.01 to possess resistance loci (Agarwal et al., 2019; Qin et al., 2012; Tseng et al., 2016; Wu et al., 2025). Three of the four studies support the beginning of Chr.01 for the region where the resistance from ‘NC94022’ resides (Agarwal et al., 2019; Qin et al., 2012; Wu et al., 2025).

The two major challenges of previous QTL studies are marker resolution and recombination frequencies. To overcome these challenges that many plant communities face, the paradigm shift to graph-based pangenome references empowers marker detection to identify recombination patterns more efficiently (Danilevicz et al., 2020; L. Guo et al., 2025; Hou et al., 2024; Liu et al., 2024; Thompson et al., 2025; Vaughn et al., 2022). A population-specific pangenome offers an unbiased reference that integrates a comprehensive set of segregating variants within a population to improve association studies (Thompson et al., 2025; Vaughn et al., 2022). Among the founders of PeanutMAGIC, ‘NC94022’ is known for TSWV resistance (Culbreath et al., 2005), however it has limited use in commercial settings and breeding programs due to poor seed traits and growth habit. While the data from PeanutMAGIC is proving exceptionally powerful, the discernment of exceedingly complex phenotypes such as pathogen resistance has not been assessed. Therefore, we hypothesize that PeanutMAGIC and pangenome are integral for the discernment of the functional variant that confers TSWV resistance to facilitate an improved and accessible germplasm for breeding.

Epidemiology of *Passalora arachidicola* (ELS) and *Nothopassalora personata* (LLS)

Early and late leaf spot (ELS, LLS) on *A. hypogaea* is caused by *Passalora arachidicola* and *Nothopassalora personata*, respectively. ELS and LLS can be affected by changes in environmental patterns across seasons and locations. Nevertheless, disease pressure each season can cause up to 70% yield loss and 100% defoliation if ELS and LLS are left untreated (Anco et al., 2020). The symptomatology of ELS and LLS are similar in many manners but there are a few key distinctions between them which allows them to be identified without the use of molecular techniques. ELS has the appearance of reddish-brown sporulation, primarily on the adaxial surface of the leaf (Shokes & Culbreath, 1997). LLS primarily have black sporulation on the abaxial surface of the leaf (Kokalis-Burelle et al., 1997). However, the discernment of these pathogens requires close inspection of the lesions, making distinct phenotyping difficult in large quantities. Therefore, observations of a field must be first calibrated to understand the presence of ELS and LLS in the study during data collection.

It is not uncommon for the same plant or even the same leaf to be infected with both ELS and LLS. Another similarity between ELS and LLS is the ability for both pathogens to adapt a polycyclic disease cycle which allows for the pathogens to reproduce exponentially, given adequate environmental conditions (Tshilenge-Lukanda et al., 2012). The disease cycle begins by conidia blowing in from past seasons debris from neighboring fields or at the margins of the field, where mycelium overwinters in the leaf litter (McDonald et al., 1985). Once the conidia attach to the leaf surface with favorable conditions of ≥95% relative humidity and temperatures between 16 and 20°C, the conidia will germinate (Shokes & Culbreath, 1997). The hyphae can either locate a natural opening like stomata or it can form an appressorium to mechanically

penetrate the leaf tissue (Shokes & Culbreath, 1997). Once in the foliar tissue there is a difference in the life cycles of ELS and LLS. ELS will kill cells and obtain nutrients from the cells through the production of phytotoxins as a necrotroph. The toxins which ELS produces to kill cells creates the yellow halo around the lesion. However, both ELS and LLS can cause chlorosis around the lesion, making the identification of a yellow halo not distinctive. LLS will feed from living cells with haustoria at first then metabolically transition to feed on dead cells as a hemibiotroph. Once the infection process has progressed, conidia are then produced and are dispersed through wind or water splash onto nearby leaves and the infection process repeats (Wadia et al., 1998). This process will repeat as the epidemic intensifies to the point where it will cause leaf defoliation, severely limiting yield (Anco et al., 2020). When defoliation occurs, especially later in the season, the leaf litter may act as an overwinter site for the pathogens as facultative saprophytes (Giordano et al., 2021).

There are multiple points throughout the life cycle of ELS and LLS that are similar and one point where they vary that may prove important for host resistance. ELS kills host cells to obtain nutrients where LLS interacts with the cell with its haustoria. Therefore, we hypothesize that there could be resistance gene(s) which are specific to ELS or LLS. Additionally, these resistance loci are likely quantitative that may also confer resistance to both ELS and LLS, or exclusively. There have been many biparental mapping populations that have been constructed to understand ELS and LLS resistance loci in cultivated peaunt (Table 1.1), many of them identifying different regions. This indicates that there are many quantitative genes that can work in unison. Therefore, ELS and LLS are quantitative traits making genomic studies for these traits difficult, however uncovering the mechanisms behind the resistance present in cultivated peanut could be added in addition to qualitative resistance from introgressions to further increase their

efficacy and longevity in application.

Interestingly one aspect of the resistance loci in *A. hypogaea* is the inconsistencies between different reports on the same biparental crosses. An example of this is where Han et al. 2018 (Han et al., 2018), Clevenger et al. 2018 (J. Clevenger et al., 2018), and Chu et al. 2019 (Chu et al., 2019) performed QTL studies on the cross 'Florida-07' X 'GP-NC WS 16'. These studies had differing resistance loci with various percent variation explained, focusing on Chr.05 and Chr.15. These inconsistencies are likely due to a variety of reasons such as small biparental populations, the difficulties of genotyping, high variance in disease pressure during different years and locations, and differences in phenotype calling and simplification of ELS and LLS differentiation. It has been shown that biparental crosses have limited strength in QTL studies due to limited crossing events and haplotype diversity (Arrones et al., 2020; Dell'Acqua et al., 2015; Scott et al., 2020). Furthermore, with the availability of higher quality genotyping methods, such as population-specific pangenome references, concerns of erroneous genotypes are reduced.

A. cardenasii Introgression Confers Peanut Root Knot Nematode Resistance

Most wild relatives of cultivated peanut are diploid, making the movement of wild alleles into the germplasm difficult (Bertioli et al., 2014, 2016, 2019). However, these alleles can offer unique traits that would improve peanut production such as disease resistance. There are two routes of introgression into cultivated peanut that have been documented, termed the tetraploid and hexaploid route. The primary route is the tetraploid route where there is an intermediate synthetic tetraploid step. An example of this approach is from C. E. Simpson at Texas A&M to introgress *A. cardenasii* alleles into the germplasm (Simpson, 1991). This approach begins with

the hybridization of an A compatible and B compatible diploid species to generate a sterile AB hybrid. This hybrid is treated with colchicine to double the genome into an AABB genome that can be crossed with cultivated peanut. This approach made the line TxAG-6 which possessed an introgression that confers resistance to peanut root knot nematodes (*Meloidogyne arenaria* race 1, RKN). This line's progeny has been used for many years with notable related cultivars including Tifguard, TifNV-High O/L, TifNV-HG, and GA-14N (Branch & Brenneman, 2015; Holbrook et al., 2008, 2017, 2023). These lines are crucial for peanut growers that tend heavily infested fields due to their minimal yield lost and to reduce the necessity of nematicides.

The PeanutMAGIC founder 'TifNV-H O/L' possesses a large introgression that occupies most of Chr.09 and was found to be the source of the population structure in the MAGIC Core (Thompson et al., 2024, 2025). This introgression is recalcitrant to recombine, however the separation of the full introgression and the causal resistant gene(s) would offer improved breeding resources for peanut breeders. Single-reference markers for the MAGIC Core identified novel recombinants of the introgression (Thompson et al., 2024), which may offer insights into the location of the resistance gene(s). Using population-specific pangenome-based markers, it was found that there are 'TifNV-H O/L' specific markers that recombine within the introgression clade and throughout the MAGIC Core (Thompson et al., 2025). This material has the potential to be used for the isolation of the resistance loci. Moreover, this could offer novel breeding material that can be used for RKN, without linkage drag.

Table 1.1: Summary of Early and Late Leaf spot QTLs.

Population	Marker Type	ELS or LLS	Major QTL Locations	Total Number of Markers	Map Density	Source
Lines in PeanutMAGIC						
SunOleic 97R X NC94022	SSR and EST	ELS	A01, A03, A05, B03, B04	248	5.7 cM/marker	(Khera et al., 2016)
		LLS	A03, B03, B05			
Tifrunner × GT-C20	SSR and EST	ELS	A03, A04, A05, A06, A07, B01, B06	418	5.3cM/marker	(Pandey et al., 2017)
		LLS	A02, A03, A04, A05, A06, A07, B04, B08			
Florida-07 x GP-NC WS16	SNP	ELS	A03, A07, B04, B10	2,753	1.34cM/marker	Han et al., 2018
		LLS	A02, A05, A07, A09, B01, B05			
Florida-07 x GP-NC WS16	SNP	LLS	A05, B03, B05	5,513	2Mb/window	Clevenger et al., 2018
Tifrunner × GT-C20	SNP	ELS	A01, A05, A06, A08, B03, B05	8,869	1.45 cM/marker	Agarwal et al., 2018
		LLS	A05, A06, B02, B03, B05, B10			
Florida-07 x GP-NC WS16	SNP and SSR	ELS	A03, A05, B03, B10	855	2.2cM/marker	Chu et al., 2019
		LLS	A03, A05, A08, B03, B05			

NAM-Tifrunner	SNP	LLS	A01, A02, B01, B02, B03, B05, B09	3,341 for linkagemap ; 11,520 for GWAS		Gangurde et al., 2024
Other Genotypes						
TAG 24 X GPBD 4	SSR	LLS	LG1, LG2, LG5, LG6, LG8, LG9, LG10, LG11, LG12, LG13	56	14 LGs: 8.25 cM/marker	Khedikar et al., 2010
TAG 24 X GPBD 4	SSR and EST	LLS	AhI, AhV, AhVIII, AhXV	209	5.124 cM/marker	Sujay et al., 2011
Zhonghua X ICGV 86699	SNP	LLS	A01, A04, A05, A06, A08, A10, B01, B04, B06, B10	1,685	0.9 cM/marker	Zhou et al., 2016
TAG 24 X GPBD 4	SSR, EST, TE	LLS	A05, B03, B10	327	6.0 cM/marker	Kolekar et al., 2016
Tanrun OL07 X Tx964117	SNP	LS	A02, B04, B06, B09, B10	1,127	5.18 cM/marker	Liang et al., 2017
TAG 24 X GPBD 4	SNP	LLS	A03	62,358	10 cM/windo w	Pandey et al., 2017a
Consensus	SSR, EST, TE, SNP	LLS	A05	5874	2.01 cM/marker	Lu et al., 2018
TAG 24 X TAG 26 X GPBD 4	SNP	LLS	A02	173,995		Shirasawa et al., 2018

GJG17 × GPBD4	SSR and EST	LLS	A03	70	11.39 cM/marker	Ahmad et al., 2020
Accession Genotypes	SNP	ELS LLS	A09, B09, B10 B09	13,382		Zhang et al., 2020

References

Agarwal, G., Clevenger, J., Kale, S. M., Wang, H., Pandey, M. K., Choudhary, D., Yuan, M., Wang, X., Culbreath, A. K., Holbrook, C. C., Liu, X., Varshney, R. K., & Guo, B. (2019). A recombination bin-map identified a major QTL for resistance to Tomato Spotted Wilt Virus in peanut (Arachis hypogaea). *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-54747-1

Agarwal, G., Clevenger, J., Pandey, M. K., Wang, H., Shasidhar, Y., Chu, Y., Fountain, J. C., Choudhary, D., Culbreath, A. K., Liu, X., Huang, G., Wang, X., Deshmukh, R., Holbrook, C. C., Bertioli, D. J., Ozias-Akins, P., Jackson, S. A., Varshney, R. K., & Guo, B. (2018). High-density genetic map using whole-genome resequencing for fine mapping and candidate gene discovery for disease resistance in peanut. *Plant Biotechnology Journal*, *16*(11), 1954–1967. https://doi.org/10.1111/pbi.12930

Alper, C. M., & Mattes, R. D. (2003). Peanut consumption improves indices of cardiovascular disease risk in healthy adults. *Journal of the American College of Nutrition*, *22*(2). https://doi.org/10.1080/07315724.2003.10719286

Anco, D. J., Thomas, J. S., Jordan, D. L., Shew, B. B., Monfort, W. S., Mehl, H. L., Small, I. M., Wright, D. L., Tillman, B. L., Dufault, N. S., Hagan, A. K., & Campbell, H. L. (2020). Peanut yield loss in the presence of defoliation caused by late or early leaf spot. *Plant Disease*, *104*(5). https://doi.org/10.1094/PDIS-11-19-2286-RE

Armstrong, J., Hickey, G., Diekhans, M., Fiddes, I. T., Novak, A. M., Deran, A., Fang, Q., Xie, D., Feng, S., Stiller, J., Genereux, D., Johnson, J., Marinescu, V. D., Alföldi, J., Harris, R. S., Lindblad-Toh, K., Haussler, D., Karlsson, E., Jarvis, E. D., … Paten, B. (2020). Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature*, *587*(7833). https://doi.org/10.1038/s41586-020-2871-y

Arrones, A., Vilanova, S., Plazas, M., Mangino, G., Pascual, L., Díez, M. J., Prohens, J., & Gramazio, P. (2020). The Dawn of the Age of Multi-Parent MAGIC Populations in Plant Breeding: Novel Powerful Next-Generation Resources for Genetic Analysis and Selection of Recombinant Elite Material. *Biology*, *9*(8), 229. https://doi.org/10.3390/biology9080229

Bertioli, D. J., Cannon, S. B., Froenicke, L., Huang, G., Farmer, A. D., Cannon, E. K. S., Liu, X., Gao, D., Clevenger, J., Dash, S., Ren, L., Moretzsohn, M. C., Shirasawa, K., Huang, W., Vidigal, B., Abernathy, B., Chu, Y., Niederhuth, C. E., Umale, P., … Ozias-Akins, P. (2016). The genome sequences of Arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut. *Nature Genetics*, *48*(4), 438–446. https://doi.org/10.1038/ng.3517

Bertioli, D. J., Jenkins, J., Clevenger, J., Dudchenko, O., Gao, D., Seijo, G., Leal-Bertioli, S. C. M., Ren, L., Farmer, A. D., Pandey, M. K., Samoluk, S. S., Abernathy, B., Agarwal, G., Ballén-Taborda, C., Cameron, C., Campbell, J., Chavarro, C., Chitikineni, A., Chu, Y., … Schmutz, J. (2019). The genome sequence of segmental allotetraploid peanut Arachis hypogaea. *Nature Genetics*, *51*(5). https://doi.org/10.1038/s41588-019-0405-z

Bertioli, D. J., Ozias-Akins, P., Chu, Y., Dantas, K. M., Santos, S. P., Gouvea, E., Guimarães, P. M., Leal-Bertioli, S. C. M., Knapp, S. J., & Moretzsohn, M. C. (2014). The use of SNP markers for linkage mapping in diploid and tetraploid peanuts. *G3: Genes, Genomes, Genetics*, *4*(1). https://doi.org/10.1534/g3.113.007617

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). https://doi.org/10.1093/bioinformatics/btm308

Branch, W. D. (2014). Registration of 'Georgia-13M' Peanut. *Journal of Plant Registrations*, *8*(3). https://doi.org/10.3198/jpr2013.11.0071crc

Branch, W. D., & Brenneman, T. B. (2015). Registration of 'Georgia-14N' Peanut. *Journal of Plant Registrations*, *9*(2). https://doi.org/10.3198/jpr2014.11.0082crc

Branch, W. D., Brown, N., & Perrera, M. A. (2022). Inheritance of Mid-Oleic Fatty Acid Ratio Seed Trait in Peanut. *Peanut Science*, *49*(2). https://doi.org/10.3146/0095-3679-492-ps22-6

Brown, N., Branch, W. D., Johnson, M., & Wallace, J. (2021). Genetic diversity assessment of Georgia peanut cultivars developed during ninety years of breeding. *The Plant Genome*, *14*(3). https://doi.org/10.1002/tpg2.20141

Chavarro, C., Chu, Y., Holbrook, C., Isleib, T., Bertioli, D., Hovav, R., Butts, C., Lamb, M., Sorensen, R., Jackson, S. A., & Ozias-Akins, P. (2019). *Pod and seed trait QTL identification to assist breeding for peanut market preferences*. https://doi.org/10.1101/738914

Chen, X., Li, H., Pandey, M. K., Yang, Q., Wang, X., Garg, V., Li, H., Chi, X., Doddamani, D., Hong, Y., Upadhyaya, H., Guo, H., Khan, A. W., Zhu, F., Zhang, X., Pan, L., Pierce, G. J., Zhou, G., Krishnamohan, K. A. V. S., … Yu, S. (2016). Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis,

and allergens. *Proceedings of the National Academy of Sciences*, *113*(24), 6785–6790. https://doi.org/10.1073/pnas.1600899113

Chen, X., Lu, Q., Liu, H., Zhang, J., Hong, Y., Lan, H., Li, H., Wang, J., Liu, H., Li, S., Pandey, M. K., Zhang, Z., Zhou, G., Yu, J., Zhang, G., Yuan, J., Li, X., Wen, S., Meng, F., … Liang, X. (2019). Sequencing of Cultivated Peanut, Arachis hypogaea, Yields Insights into Genome Evolution and Oil Improvement. *Molecular Plant*, *12*(7). https://doi.org/10.1016/j.molp.2019.03.005

Chu, Y., Chee, P., Culbreath, A., Isleib, T. G., Holbrook, C. C., & Ozias-Akins, P. (2019). Major QTLs for Resistance to Early and Late Leaf Spot Diseases Are Identified on Chromosomes 3 and 5 in Peanut (Arachis hypogaea). *Frontiers in Plant Science*, *10*. https://doi.org/10.3389/fpls.2019.00883

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., Ramos, L., Holbrook, C. C., & Ozias-Akins, P. (2007). Frequency of a loss-of-function mutation in oleoyl-PC desaturase (ahFAD2A) in the mini-core of the U.S. peanut germplasm collection. *Crop Science*, *47*(6). https://doi.org/10.2135/cropsci2007.02.0117

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*, *4*(2), 110–117. https://doi.org/10.3835/plantgenome2011.01.0001

Clevenger, J., Chavarro, C., Pearl, S. A., Ozias-Akins, P., & Jackson, S. A. (2015). Single nucleotide polymorphism identification in polyploids: A review, example, and recommendations. In *Molecular Plant* (Vol. 8, Issue 6). https://doi.org/10.1016/j.molp.2015.02.002

Clevenger, J., Chu, Y., Arrais Guimaraes, L., Maia, T., Bertioli, D., Leal-Bertioli, S., Timper, P., Holbrook, C. C., & Ozias-Akins, P. (2017). Gene expression profiling describes the genetic regulation of Meloidogyne arenaria resistance in Arachis hypogaea and reveals a candidate gene for resistance. *Scientific Reports*, *7*(1), 1317. https://doi.org/10.1038/s41598-017-00971-6

Clevenger, J., Chu, Y., Chavarro, C., Botton, S., Culbreath, A., Isleib, T. G., Holbrook, C. C., & Ozias-Akins, P. (2018). Mapping late leaf spot resistance in peanut (Arachis hypogaea)

using QTL-seq reveals markers for marker-assisted selection. *Frontiers in Plant Science*, *9*. https://doi.org/10.3389/fpls.2018.00083

Clevenger, J. P., & Ozias-Akins, P. (2015). SWEEP: A tool for filtering high-quality SNPs in polyploid crops. *G3: Genes, Genomes, Genetics*, *5*(9). https://doi.org/10.1534/g3.115.019703

Culbreath, A. K., Gorbet, D. W., Martinez-Ochoa, N., Holbrook, C. C., Todd, J. W., Isleib, T. G., & Tillman, B. (2005). High Levels of Field Resistance to Tomato spotted wilt virus in Peanut Breeding Lines Derived from hypogaea and hirsuta Botanical Varieties. *Peanut Science*, *32*(1). https://doi.org/10.3146/0095-3679(2005)32[20:hlofrt]2.0.co;2

Danilevicz, M. F., Tay Fernandez, C. G., Marsh, J. I., Bayer, P. E., & Edwards, D. (2020). Plant pangenomics: approaches, applications and advancements. In *Current Opinion in Plant Biology* (Vol. 54). https://doi.org/10.1016/j.pbi.2019.12.005

Davis, J. P., Price, K. M., Dean, L. L., Sweigart, D. S., Cottonaro, J. M., & Sanders, T. H. (2016). Peanut Oil Stability and Physical Properties Across a Range of Industrially Relevant Oleic Acid/Linoleic Acid Ratios. *Peanut Science*, *43*(1). https://doi.org/10.3146/0095-3679-43.1.1

Dell'Acqua, M., Gatti, D. M., Pea, G., Cattonaro, F., Coppens, F., Magris, G., Hlaing, A. L., Aung, H. H., Nelissen, H., Baute, J., Frascaroli, E., Churchill, G. A., Inzé, D., Morgante, M., & Pè, M. E. (2015). Genetic properties of the MAGIC maize population: a new platform for high definition QTL mapping in Zea mays. *Genome Biology*, *16*(1), 167. https://doi.org/10.1186/s13059-015-0716-z

Gangurde, S. S., Thompson, E., Yaduru, S., Wang, H., Fountain, J. C., Chu, Y., Ozias-Akins, P., Isleib, T. G., Holbrook, C., Dutta, B., Culbreath, A. K., Pandey, M. K., & Guo, B. (2024). Linkage Mapping and Genome-Wide Association Study Identified Two Peanut Late Leaf Spot Resistance Loci, *PLLSR* -1 and *PLLSR* -2, Using Nested Association Mapping. *Phytopathology*®, *114*(6), 1346–1355. https://doi.org/10.1094/PHYTO-04-23-0143-R

Gangurde, S. S., Wang, H., Yaduru, S., Pandey, M. K., Fountain, J. C., Chu, Y., Isleib, T., Holbrook, C. C., Xavier, A., Culbreath, A. K., Ozias-Akins, P., Varshney, R. K., & Guo, B. (2020). Nested-association mapping (NAM)-based genetic dissection uncovers candidate genes for seed and pod weights in peanut (*Arachis hypogaea*). *Plant Biotechnology Journal*, *18*(6), 1457–1471. https://doi.org/10.1111/pbi.13311

Garrison, E., Sirén, J., Novak, A. M., Hickey, G., Eizenga, J. M., Dawson, E. T., Jones, W., Garg, S., Markello, C., Lin, M. F., Paten, B., & Durbin, R. (2018). Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nature Biotechnology*, *36*(9). https://doi.org/10.1038/nbt.4227

Gorbet, D. W., & Knauft, D. A. (2000). Registration of 'SunOleic 97R' Peanut. *Crop Science*, *40*(4). https://doi.org/10.2135/cropsci2000.0032rcv

Gorbet, D. W., & Tillman, B. L. (2009). Registration of 'Florida-07' Peanut. *Journal of Plant Registrations*, *3*(1). https://doi.org/10.3198/jpr2008.05.0276crc

Guo, B., Chen, C., Chu, Y., Holbrook, C. C., Ozias-Akins, P., & Stalker, H. T. (2012). Advances in Genetics and Genomics for Sustainable Peanut Production. In *Sustainable Agriculture and New Biotechnologies* (pp. 362–389). CRC Press. https://doi.org/10.1201/b10977-20

Guo, B., Khera, P., Wang, H., Peng, Z., Sudini, H., Wang, X., Osiru, M., Chen, J., Vadez, V., Yuan, M., Wang, C. T., Zhang, X., Waliyar, F., Wang, J., & Varshney, R. K. (2016). Annotation of Trait Loci on Integrated Genetic Maps of Arachis Species. In *Peanuts* (pp. 163–207). Elsevier. https://doi.org/10.1016/B978-1-63067-038-2.00006-X

Guo, L., Wang, X., Ayhan, D. H., Rhaman, M. S., Yan, M., Jiang, J., Wang, D., Zheng, W., Mei, J., Ji, W., Jiao, J., Chen, S., Sun, J., Yi, S., Meng, D., Wang, J., Bhuiyan, M. N., Qin, G., Guo, L., … Ye, W. (2025). Super pangenome of Vitis empowers identification of downy mildew resistance genes for grapevine improvement. *Nature Genetics*, *57*(3), 741–753. https://doi.org/10.1038/s41588-025-02111-7

Han, S., Yuan, M., Clevenger, J. P., Li, C., Hagan, A., Zhang, X., Chen, C., & He, G. (2018). A snp-based linkage map revealed QTLs for resistance to early and late leaf spot diseases in peanut (Arachis hypogaea L.). *Frontiers in Plant Science*, *9*. https://doi.org/10.3389/fpls.2018.01012

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). https://doi.org/10.1007/s00122-004-1756-z

Hickey, G., Heller, D., Monlong, J., Sibbesen, J. A., Sirén, J., Eizenga, J., Dawson, E. T., Garrison, E., Novak, A. M., & Paten, B. (2020). Genotyping structural variants in pangenome graphs using the vg toolkit. *Genome Biology*, *21*(1). https://doi.org/10.1186/s13059-020-1941-7

Hickey, G., Monlong, J., Ebler, J., Novak, A. M., Eizenga, J. M., Gao, Y., Abel, H. J., Antonacci-Fulton, L. L., Asri, M., Baid, G., Baker, C. A., Belyaeva, A., Billis, K., Bourque, G., Buonaiuto, S., Carroll, A., Chaisson, M. J. P., Chang, P. C., Chang, X. H., … Paten, B. (2023). Pangenome graph construction from genome alignments with Minigraph-Cactus. *Nature Biotechnology*. https://doi.org/10.1038/s41587-023-01793-w

Holbrook, C. C. (1983). A Technique for Screening Peanut for Resistance to Meloidogyne arenaria . *Plant Disease*, *67*(9). https://doi.org/10.1094/pd-67-957

Holbrook, C. C., & Culbreath, A. K. (2007). Registration of ‘Tifrunner’ Peanut. *Journal of Plant Registrations*, *1*(2), 124–124. https://doi.org/10.3198/jpr2006.09.0575crc

Holbrook, C. C., Ozias-Akins, P., Chu, Y., Brenneman, T. B., & Culbreath, A. K. (2023). Registration of ‘TifNV-HG’ peanut. *Journal of Plant Registrations*, *17*(3). https://doi.org/10.1002/plr2.20295

Holbrook, C. C., Ozias-Akins, P., Chu, Y., Culbreath, A. K., Kvien, C. K., & Brenneman, T. B. (2017). Registration of ‘TifNV-High O/L’ Peanut. *Journal of Plant Registrations*, *11*(3). https://doi.org/10.3198/jpr2016.10.0059crc

Holbrook, C. C., Timper, P., Culbreath, A. K., & Kvien, C. K. (2008). Registration of ‘Tifguard’ Peanut. *Journal of Plant Registrations*, *2*(2). https://doi.org/10.3198/jpr2007.12.0662crc

Hong, Y., Chen, X., Liang, X., Liu, H., Zhou, G., Li, S., Wen, S., Holbrook, C. C., & Guo, B. (2010). A SSR-based composite genetic linkage map for the cultivated peanut (Arachis hypogaea L.) genome. *BMC Plant Biology*, *10*(1), 17. https://doi.org/10.1186/1471-2229-10-17

Hou, Y., Gan, J., Fan, Z., Sun, L., Garg, V., Wang, Y., Li, S., Bao, P., Cao, B., Varshney, R. K., & Zhao, H. (2024). Haplotype-based pangenomes reveal genetic variations and climate adaptations in moso bamboo populations. *Nature Communications*, *15*(1), 8085. https://doi.org/10.1038/s41467-024-52376-5

Husted, L. (1936). Cytological Studies an the Peanut, <i>Arachis</i>. II. *CYTOLOGIA*, *7*(3), 396–423. https://doi.org/10.1508/cytologia.7.396

Isleib, T. G. ., Holbrook, C. C., & Gorbet, D. W. (2001). Use of Plant Introductions in Peanut Cultivar Development. *Peanut Science*, *28*(2), 96–113. https://doi.org/10.3146/i0095-3679-28-2-11

Jones, J. B., Provost, M., Keaver, L., Breen, C., Ludy, M.-J., & Mattes, R. D. (2014). A randomized trial on the effects of flavorings on the health benefits of daily peanut consumption. *The American Journal of Clinical Nutrition*, *99*(3), 490–496. https://doi.org/10.3945/ajcn.113.069401

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873). https://doi.org/10.1038/s41586-021-03819-2

Jung, S., Powell, G., Moore, K., & Abbott, A. (2000). The high oleate trait in the cultivated peanut [Arachis hypogaea L.]. II. Molecular basis and genetics of the trait. *Molecular and General Genetics MGG*, *263*(5), 806–811. https://doi.org/10.1007/s004380000243

Kassambara, A. (2020). ggpubr: “ggplot2” based publication ready plots. R package version 0.2. *Https://CRAN.R-Project.Org/Package=ggpubr*.

Khera, P., Pandey, M. K., Wang, H., Feng, S., Qiao, L., Culbreath, A. K., Kale, S., Wang, J., Holbrook, C. C., Zhuang, W., Varshney, R. K., & Guo, B. (2016). Mapping Quantitative Trait Loci of Resistance to Tomato Spotted Wilt Virus and Leaf Spots in a Recombinant Inbred Line Population of Peanut (Arachis hypogaea L.) from SunOleic 97R and NC94022. *PLOS ONE*, *11*(7), e0158452. https://doi.org/10.1371/journal.pone.0158452

Kokot, M., Dlugosz, M., & Deorowicz, S. (2017). KMC 3: counting and manipulating k-mer statistics. *Bioinformatics (Oxford, England)*, *33*(17). https://doi.org/10.1093/bioinformatics/btx304

Korani, W., O’Connor, D., Chu, Y., Chavarro, C., Ballen, C., Guo, B., Ozias-Akins, P., Wright, G., & Clevenger, J. (2021). De novo QTL-seq identifies loci linked to blanchability in peanut (Arachis hypogaea) and refines previously identified QTL with low coverage sequence. *Agronomy*, *11*(11). https://doi.org/10.3390/agronomy11112201

Kris-Etherton, P. M., Hu, F. B., Ros, E., & Sabaté, J. (2008). The role of tree nuts and peanuts in the prevention of coronary heart disease: Multiple potential mechanisms. *Journal of Nutrition*, *138*(9). https://doi.org/10.1093/jn/138.9.1746s

Leastro, M. O., De Oliveira, A. S., Pallás, V., Sánchez-Navarro, J. A., Kormelink, R., & Resende, R. O. (2017). The NSm proteins of phylogenetically related tospoviruses trigger Sw-5b–mediated resistance dissociated of their cell-to-cell movement function. *Virus Research*, *240*, 25–34. https://doi.org/10.1016/j.virusres.2017.07.019

Li, F., Wang, J., Ma, C., Zhao, Y., Wang, Y., Hasi, A., & Qi, Z. (2013). Glutamate Receptor-Like Channel3.3 Is Involved in Mediating Glutathione-Triggered Cytosolic Calcium Transients, Transcriptional Changes, and Innate Immunity Responses in Arabidopsis. *PLANT PHYSIOLOGY*, *162*(3), 1497–1509. https://doi.org/10.1104/pp.113.217208

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H. (2021). New strategies to improve minimap2 alignment accuracy. *Bioinformatics*, *37*(23). https://doi.org/10.1093/bioinformatics/btab705

Liang, X. Q., Holbrook, C. C., Lynch, R. E., & Guo, B. Z. (2005). β-1,3-Glucanase Activity in Peanut Seed (*Arachis hypogaea*) is Induced by Inoculation with *Aspergillus flavus* and Copurifies with a Conglutin-Like Protein. *Phytopathology®*, *95*(5), 506–511. https://doi.org/10.1094/PHYTO-95-0506

Liao, W. W., Asri, M., Ebler, J., Doerr, D., Haukness, M., Hickey, G., Lu, S., Lucas, J. K., Monlong, J., Abel, H. J., Buonaiuto, S., Chang, X. H., Cheng, H., Chu, J., Colonna, V.,

Eizenga, J. M., Feng, X., Fischer, C., Fulton, R. S., … Paten, B. (2023). A draft human pangenome reference. *Nature*, *617*(7960). https://doi.org/10.1038/s41586-023-05896-x

Liu, Z., Wang, N., Su, Y., Long, Q., Peng, Y., Shangguan, L., Zhang, F., Cao, S., Wang, X., Ge, M., Xue, H., Ma, Z., Liu, W., Xu, X., Li, C., Cao, X., Ahmad, B., Su, X., Liu, Y., … Zhou, Y. (2024). Grapevine pangenome facilitates trait genetics and genomic breeding. *Nature Genetics*, *56*(12), 2804–2814. https://doi.org/10.1038/s41588-024-01967-5

López, Y., Nadaf, H. L., Smith, O. D., Connell, J. P., Reddy, A. S., & Fritz, A. K. (2000). Isolation and characterization of the Δ12-fatty acid desaturase in peanut (Arachis hypogaea L.) and search for polymorphisms for the high oleate trait in Spanish market-type lines. *Theoretical and Applied Genetics*, *101*(7). https://doi.org/10.1007/s001220051589

Luo, M., Dang, P., Guo, B. Z., He, G., Holbrook, C. C., Bausher, M. G., & Lee, R. D. (2005). Generation of Expressed Sequence Tags (ESTs) for Gene Discovery and Marker Development in Cultivated Peanut. *Crop Science*, *45*(1), 346–353. https://doi.org/10.2135/cropsci2005.0346

Luo, Z., Cui, R., Chavarro, C., Tseng, Y.-C., Zhou, H., Peng, Z., Chu, Y., Yang, X., Lopez, Y., Tillman, B., Dufault, N., Brenneman, T., Isleib, T. G., Holbrook, C., Ozias-Akins, P., & Wang, J. (2020). Mapping quantitative trait loci (QTLs) and estimating the epistasis controlling stem rot resistance in cultivated peanut (Arachis hypogaea). *Theoretical and Applied Genetics*, *133*(4), 1201–1212. https://doi.org/10.1007/s00122-020-03542-y

McDonald, D., Subrahmanyam, P., Gibbons, R. W., & Smith, D. H. (1985). *Early and Late Leaf Spots of Groundnut*.

Moury, B., Palloix, A., Gebre Selassie, K., & Marchoux, G. (1997). Hypersensitive resistance to tomato spotted wilt virus in three Capsicum chinense accessions is controlled by a single gene and is overcome by virulent strains. *Euphytica*, *94*(1), 45–52. https://doi.org/10.1023/A:1002997522379

Nguyen, C. T., Kurenda, A., Stolz, S., Chételat, A., & Farmer, E. E. (2018). Identification of cell populations necessary for leaf-to-leaf electrical signaling in a wounded plant. *Proceedings of the National Academy of Sciences*, *115*(40), 10178–10183. https://doi.org/10.1073/pnas.1807049115

Pandey, M. K., Wang, H., Khera, P., Vishwakarma, M. K., Kale, S. M., Culbreath, A. K., Holbrook, C. C., Wang, X., Varshney, R. K., & Guo, B. (2017). Genetic Dissection of Novel QTLs for Resistance to Leaf Spots and Tomato Spotted Wilt Virus in Peanut (Arachis hypogaea L.). *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.00025

Pandey, M. K., Wang, M. L., Qiao, L., Feng, S., Khera, P., Wang, H., Tonnis, B., Barkley, N. A., Wang, J., Holbrook, C. C., Culbreath, A. K., Varshney, R. K., & 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). https://doi.org/10.1186/s12863-014-0133-4

Qin, H., Feng, S., Chen, C., Guo, Y., Knapp, S., Culbreath, A., He, G., Wang, M. L., Zhang, X., Holbrook, C. C., Ozias-Akins, P., & Guo, B. (2012). An integrated genetic linkage map of cultivated peanut (Arachis hypogaea L.) constructed from two RIL populations. *Theoretical and Applied Genetics*, *124*(4), 653–664. https://doi.org/10.1007/s00122-011-1737-y

R Core Team. (2014). R Core Team (2014). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria. URL Http://Www.R-Project.Org/*.

R Core Team. (2021). R core team (2021). In *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www. R-project. org*.

Roy, J., Soler-Garzón, A., Miklas, P. N., Lee, R., Clevenger, J., Myers, Z., Korani, W., & McClean, P. E. (2023). Integrating de novo QTL-seq and linkage mapping to identify quantitative trait loci conditioning physiological resistance and avoidance to white mold disease in dry bean. *The Plant Genome*, *16*(4). https://doi.org/10.1002/tpg2.20380

Scott, M. F., Ladejobi, O., Amer, S., Bentley, A. R., Biernaskie, J., Boden, S. A., Clark, M., Dell'Acqua, M., Dixon, L. E., Filippi, C. V., Fradgley, N., Gardner, K. A., Mackay, I. J., O'Sullivan, D., Percival-Alwyn, L., Roorkiwal, M., Singh, R. K., Thudi, M., Varshney, R. K., … Mott, R. (2020). Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. In *Heredity* (Vol. 125, Issue 6). https://doi.org/10.1038/s41437-020-0336-6

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

Shokes, F. M., & Culbreath, A. K. (1997). *Early and late leaf spots* (N. Kokalis-Burelle, D. M. Porter, K. R. Rodriguez, D. H. Smith, & P. Subrahmanyam, Eds.; 2nd ed.). American Phytopathological Society.

Simpson, C. E. (1991). Pathways for Introgression of Pest Resistance into Arachis hypogaea L.1 . *Peanut Science*, *18*(1). https://doi.org/10.3146/i0095-3679-18-1-8

Sirén, J., Eskandar, P., Ungaro, M. T., Hickey, G., Eizenga, J. M., Novak, A. M., Chang, X., Chang, P.-C., Kolmogorov, M., Carroll, A., Monlong, J., & Paten, B. (2024). Personalized

pangenome references. *Nature Methods*, *21*(11), 2017–2023. https://doi.org/10.1038/s41592-024-02407-2

Sirén, J., Monlong, J., Chang, X., Novak, A. M., Eizenga, J. M., Markello, C., Sibbesen, J. A., Hickey, G., Chang, P.-C., Carroll, A., Gupta, N., Gabriel, S., Blackwell, T. W., Ratan, A., Taylor, K. D., Rich, S. S., Rotter, J. I., Haussler, D., Garrison, E., & Paten, B. (2021). Pangenomics enables genotyping of known structural variants in 5202 diverse genomes. *Science*, *374*(6574). https://doi.org/10.1126/science.abg8871

Spassova, M. I., Prins, T. W., Folkertsma, R. T., Klein-Lankhorst, R. M., Hille, J., Goldbach, R. W., & Prins, M. (2001). The tomato gene Sw5 is a member of the coiled coil, nucleotide binding, leucine-rich repeat class of plant resistance genes and confers resistance to TSWV in tobacco. *Molecular Breeding*, *7*(2), 151–161. https://doi.org/10.1023/A:1011363119763

Stange, M., Utz, H. F., Schrag, T. A., Melchinger, A. E., & Würschum, T. (2013). High-density genotyping: an overkill for QTL mapping? Lessons learned from a case study in maize and simulations. *Theoretical and Applied Genetics*, *126*(10), 2563–2574. https://doi.org/10.1007/s00122-013-2155-0

Sun, C., Jin, L., Cai, Y., Huang, Y., Zheng, X., & Yu, T. (2019). l-Glutamate treatment enhances disease resistance of tomato fruit by inducing the expression of glutamate receptors and the accumulation of amino acids. *Food Chemistry*, *293*, 263–270. https://doi.org/10.1016/j.foodchem.2019.04.113

Tallury, S. P., Isleib, T. G., Copeland, S. C., Rosas-Anderson, P., Balota, M., Singh, D., & Stalker, H. T. (2014). Registration of Two Multiple Disease-Resistant Peanut Germplasm Lines Derived from *Arachis cardenasii* Krapov. & W.C. Gregory, GKP 10017. *Journal of Plant Registrations*, *8*(1), 86–89. https://doi.org/10.3198/jpr2013.04.0017crg

Thompson, E., Korani, W., Wu, D., Garg, V., Tonnis, B., Wang, M., Holbrook, C. C., Ozias-Akins, P., Culbreath, A. K., Varshney, R. K., Guo, B., & Clevenger, J. (2025). PeanutMAGIC multiparent population-specific pangenome unveils a novel third FAD2 gene and solves the mystery of mid-oleic fatty acid in peanut. *Nature Communications, In Review*.

Thompson, E., Wang, H., Korani, W., Fountain, J. C., Culbreath, A. K., Holbrook, C. C., Clevenger, J. P., & Guo, B. (2024). Genetic and genomic characterization of a multiparent advanced generation intercross (MAGIC) population of peanut (*Arachis hypogaea* L.). *Crop Science*, *65*(1). https://doi.org/10.1002/csc2.21402

Tseng, Y.-C., Tillman, B. L., Peng, Z., & Wang, J. (2016). Identification of major QTLs underlying tomato spotted wilt virus resistance in peanut cultivar Florida-EPTM ‘113.’ *BMC Genetics*, *17*(1), 128. https://doi.org/10.1186/s12863-016-0435-9

Tshilenge-Lukanda, L., Nkongolo, K. K. C., Kalonji-Mbuyi, A., & Kizungu, R. V. (2012). Epidemiology of the Groundnut (<i>Arachis hypogaea</i> L.) Leaf Spot Disease: Genetic Analysis and Developmental Cycles. *American Journal of Plant Sciences*, *03*(05). https://doi.org/10.4236/ajps.2012.35070

Vaughn, J. N., Branham, S. E., Abernathy, B., Hulse-Kemp, A. M., Rivers, A. R., Levi, A., & Wechter, W. P. (2022). Graph-based pangenomics maximizes genotyping density and reveals structural impacts on fungal resistance in melon. *Nature Communications*, *13*(1), 7897. https://doi.org/10.1038/s41467-022-35621-7

Videira, S. I. R., Groenewald, J. Z., Nakashima, C., Braun, U., Barreto, R. W., de Wit, P. J. G. M., & Crous, P. W. (2017). Mycosphaerellaceae – Chaos or clarity? *Studies in Mycology*, *87*. https://doi.org/10.1016/j.simyco.2017.09.003

Wadia, K. D. R., McCartney, H. A., & Butler, D. R. (1998). Dispersal of Passalora personata conidia from groundnut by wind and rain. *Mycological Research*, *102*(3). https://doi.org/10.1017/S0953756297004887

Wang, H., Lei, Y., Wan, L., Yan, L., Lv, J., Dai, X., Ren, X., Guo, W., Jiang, H., & Liao, B. (2016). Comparative transcript profiling of resistant and susceptible peanut post-harvest seeds in response to aflatoxin production by Aspergillus flavus. *BMC Plant Biology*, *16*(1), 54. https://doi.org/10.1186/s12870-016-0738-z

Wickham, H. (2011). Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. In *Wiley Interdisciplinary Reviews: Computational Statistics* (Vol. 3, Issue 2).

Wu, D., Zhao, C., Korani, W., Thompson, E. A., Wang, H., Agarwal, G., Fountain, J. C., Culbreath, A., Holbrook, C. C., Wang, X., Clevenger, J. P., & Guo, B. (2025). High-resolution genetic and physical mapping reveals a peanut spotted wilt disease resistance locus, PSWDR-1, to Tomato spotted wilt virus (TSWV), within a recombination cold-spot on chromosome A01. *BMC Genomics*, *26*(1), 224. https://doi.org/10.1186/s12864-025-11366-7

Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M. K., Zhang, C., Chang, W.-C., Zhang, L., Zhang, X., Tang, R., Garg, V., Wang, X., Tang, H., Chow, C.-N., Wang, J., Deng, Y., Wang, D., Khan, A. W., Yang, Q., … Varshney, R. K. (2019). The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nature Genetics*, *51*(5), 865–876. https://doi.org/10.1038/s41588-019-0402-2

CHAPTER 2

GENETIC AND GENOMIC CHARACTERIZATION OF A MULTIPARENT ADVANCED GENERATION INTERCROSS (MAGIC) POPULATION OF PEANUT (ARACHIS HYPOGAEA L.)

[1] Thompson, E., Wang, H., Korani, W., Fountain, J. C., Culbreath, A. K., Holbrook, C. C., Clevenger, J. P., & Guo, B. (2024). Genetic and genomic characterization of a multiparent advanced generation intercross (MAGIC) population of peanut (*Arachis hypogaea* L.*). Crop Science*, 65(1). https://doi.org/10.1002/csc2.21402

Abstract

Multiparent advanced generation inter-cross (MAGIC) populations are a new genetic resource for high-resolution mapping of quantitative traits and as a source of new germplasm or improved cultivars for breeding due to the high level of recombination events in the population. Here, we have developed an eight-founder MAGIC population for peanut (PeanutMAGIC). Eight diverse founders were inter-crossed using a simple funnel mating design to ensure that the MAGIC population would possess equal representation from each founder. This was followed by advancement using small family plot and single-seed descent, resulting in 3,187 $F_{2:7}$ recombinant inbred lines (RILs). The objective of this study was to introduce this PeanutMAGIC as a new resource for genetic and genomic studies. We randomly selected a smaller subset of 310 RILs (MAGIC Core) from PeanutMAGIC and conducted genotyping using whole genome resequencing and phenotyping over two growing seasons for seed and pod traits. Whole genome characterization of the MAGIC Core demonstrated that PeanutMAGIC harbors a balanced and evenly differentiated mosaic of genomic blocks from the eight founders, providing unique recombination events for high-resolution mapping of quantitative traits. Using two-year phenotypic data, we showed that PeanutMAGIC can improve genetic mapping power of a spectrum of qualitative, like seed coat color, to quantitative traits such as pod weight, seed weight, shelling percentage, pod constriction, and pod reticulation. These findings show that the PeanutMAGIC population can be used by the peanut research community as a new resource for genetic and genomic studies and for cultivar improvement.

Introduction

Cultivated peanut (*Arachis hypogaea* L.) is an important legume crop grown worldwide. However, cultivated peanut has a narrow genetic base with limited genetic diversity because of its origin from a recent hybridization and genome duplication of two diploid species, *A. duranensis* (A-genome) and *A. ipaënsis* (B-genome) (Bertioli et al., 2019). Cultivated peanut genomic architecture is allotetraploid (2n = 4x = 40) with two sets of chromosomes (Chr.) pairs (AABB; AA = Chr. 01-10, BB = Chr. 11-20), where wild diploid relatives (2n = 2x = 20) have mostly AA- or BB-type genomes (Bertioli et al., 2016, 2019), restricting the movement of wild alleles into the cultivated peanut germplasm. As early as 1936, Husted (1936) hypothesized that cultivated peanut developed from a single hybridization of two diploid progenitors, which was supported by recent genomic technologies (Seijo et al., 2004; Bertioli et al., 2016, 2019; Zhuang et al., 2019). The fact of low genetic diversity of cultivated peanut has limited genetic and genomic studies in peanut community (Guo et al., 2012, 2016) before the completion of peanut genome sequencing project (Bertioli et al., 2016, 2019), despite the continual efforts of peanut researchers like breeders (Isleib et al., 2001; Brown et al., 2021) to broad genetic diversity by bringing new germplasm into breeding programs. These efforts have been driven by the need for disease resistance introduced through germplasm and interspecific hybridization with diploid relatives (Isleib et al., 2001).

For decades, biparental mapping populations have been the “standard” genomic and breeding approach for plant species (Scott et al., 2020). However, the limited meiotic recombination, diversity, and population size associated with this approach have hindered these approaches for the peanut research community. Most peanut genetic mapping studies have used biparental recombinant inbred line (RIL) populations, such as the “S” and the “T” RIL

populations (Qin et al., 2012). These studies also used simple sequence repeat (SSR) markers (Luo et al., 2005; Hong et al., 2010), resulting in lower resolution in QTL prediction (Khera et al., 2016; Pandey et al., 2017). Recent studies by Agarwal et al. (2018, 2019), using the same "S" and "T" RIL populations, demonstrated the power of genotyping by whole genome resequencing (WGRS), which increased the maker density and the ability to track recombination events compared to SSR marker approaches.

In the last decade, the worldwide peanut community has witnessed the progress made in genomic research through the availability of high-quality reference genomes for both diploid and tetraploid peanuts (Bertioli et al., 2016, 2019; Chen et al., 2016, 2019; Zhuang et al., 2019). As cost of sequencing decreases and accuracy in marker selection increases, particularly with polyploids like peanuts (Clevenger et al., 2015a, 2015b), marker saturation is no longer a limiting factor in genetic mapping of traits of interest (Stange et al., 2013; Agarwal et al., 2019; Korani et al., 2021). As a result, there is a growing need for populations with greater genetic diversity and recombination density. Recently, the availability of genomic tools designed for duplicated understudied genomes (Korani et al., 2021; Roy et al., 2023) has facilitated the dissection of complex populations, making the construction and observation of these populations accessible.

Multiparent populations are the converging point between biparental populations and diversity panels. Biparental populations have the advantage of being easy to synthesize and having known pedigrees. However, biparental populations suffer from a lack of diversity and limited recombination (Arrones et al., 2020). Diversity panels are the opposite, where there is high genetic diversity and historical accumulation of recombination events in the panel, but they are limited by unknown pedigrees and parental information (Dell'Acqua et al., 2015; Arrones et

al., 2020). Multiparent populations bridge the gap, allowing for the combination of several diverse lines into one population, thus, increasing recombination and diversity while preserving the pedigree information of the population (Dell'Acqua et al., 2015). There are several multiparental population structures utilized in plant breeding such as Nested-Associated Mapping (NAM) and Multiparent Advanced Generation Inter-Cross (MAGIC) (Scott et al., 2020). NAM is a breeding design where a common parent is crossed to an array of other unique parental genotypes. There are limitations to NAM population structures, including a focus on dissecting the common parent's genotype through recombination with a single other founder in each panel. While peanut NAM populations have been constructed (Holbrook et al., 2013), the individual NAM families have been primarily used in a biparental population fashion (Chu et al., 2019a, 2019b; Chavarro et al., 2020; Luo et al., 2020) because of the logistical challenges of working with large number of multiparental populations. Gangurde et al. (2019, 2024) reported the power of the NAM approach in mapping pod and seed traits and late leaf spot resistance using a subset of these peanut NAM populations.

In contrast, MAGIC populations facilitate a mosaic of all founders equally within the population through multiple crossing generations and distinct haplotypes combinations, resulting in the capability to dissect complex genetic traits. This breeding design supports plant breeding by broadening the genetic base and pyramiding traits into lines that can be used in cultivar development. During the inter-crossing generations, the meiotic recombination events are the essential process that rearranges alleles and produces new combinations of genes for breeding lines. Here we introduce and characterize a peanut MAGIC population, named PeanutMAGIC, which was derived from eight diverse peanut cultivars and breeding lines through a simple funnel design (Supplementary Figure 2.1). PeanutMAGIC is a new mapping population with

high genetic diversity and fine linkage disequilibrium (LD) blocks, which will allow for the dissection of complex traits and offer an additional resource for peanut germplasm improvement. We selected a subset of 310 RILs from PeanutMAGIC (3,187 RILs), called MAGIC Core, for easy to use in evaluating and maintenance (Holbrook et al., 1993; Holbrook and Dong, 2005), conducted two-year phenotyping study, and performed genotyping by whole genome resequencing to characterize the PeanutMAGIC. While not publicly released, small seed lots of the PeanutMAGIC or the MAGIC Core are available from the corresponding author for research and breeding purposes, free of charge upon request.

Materials and Methods

Development of peanut MAGIC population (PeanutMAGIC)[1]

A multiparental advanced generation inter-cross (MAGIC) population for cultivated peanut was constructed with eight diverse peanut cultivars and breeding lines using a simple "funnel" crossing scheme (Supplementary Figure 2.1). The eight founders (G_0) are Tifrunner (Holbrook et al., 2007), GT-C20 (Liang et al., 2005), NC94022 (Culbreath et al., 2005), SunOleic 97R (Gorbet and Knauft, 2000), Georgia-13M (GA-13M) (Branch, 2014), TifNV-High O/L (TifNV-H O/L) (Holbrook et al., 2017), Florida-07 (FLA-07) (Gorbet and Tillman, 2009), and GP-NC WS16 (Tallury et al., 2014) and were denoted as A to H, respectively. The founders were crossed in a simple funnel design and crosses were made reciprocally for maximizing the numbe[1]r of pollinations made, which results in recombinant inbred lines (RILs) with different maternal parents. In each crossing generation the true hybrids were identified through PCR

[1] The population was developed by Hui Wang under the supervision of Baozhu Guo

(hereafter all numbers reported here were numbers of true hybrids) (Supplementary Figure 2.1). The initial four 2-way crosses produced 2-way hybrids (G_1) of 19 AB, 17 CD, 27 EF, and 26 GH. The G_1 hybrids were used for two 4-way crosses, creating 210 ABCD and 172 EFGH 4-way hybrid seeds (G_2). The increased number of 4-way hybrid seeds produced were critical for the 8-way crossing because each G_2 hybrid seed has a unique combination of alleles between two donors on each chromatid amplifying the novel combinations of genetic material during the 8-way cross.

For the 8-way crosses, we successfully made 150 specific pairs of one-to-one crosses, using one flower from a specific 4-way hybrid plant (ABCD) to cross to a flower of a specific corresponding plant from the other 4-way hybrids (EFGH), or *vice versa* (reciprocal), resulting in 950 8-way hybrid F_1 seeds. We harvested the 8-way hybrid F_1 seeds individually and coded with two numbers for each seed, the pair number and the F_1 seed number produced by this pair. For example, in Figure 2.1 "39-9" indicates that this F_1 plant was the 9^{th} hybrid F_1 seed produced by the 8-way pair No. 39. The 950 8-way hybrid F_1 seeds were planted in the summer of 2018 in Tifton, Georgia (Figure 2.1). To maximize the production of F_2 seeds, we planted the 8-way F_1 seeds individually with a larger space and alternating with a border plant of Georgia-06G (Branch, 2007), a popular commercial variety. The border plants were later removed to provide additional room for the F_1 plant. In total 3,575 F_2 seeds were harvested from all F_1 plants. All $F_{2:3}$ families were advanced through bulked family small plot seeds at each generation until the F_5 generation. In the summer of 2020, individual F_5 plants were selected at random and harvested from each F_5 family-plot as single seed descend (SSD) to generate $F_{5:6}$ families. To ensure the equal representation of all 8-way crosses, we selected the same number of $F_{2:6}$ families from each 8-way cross. In the summer of 2021, $F_{6:7}$ seeds were increased and harvested, creating

3,187 $F_{2:7}$ PeanutMAGIC RILs. For this study, a subset of 310 RILs were randomly selected to create the MAGIC Core for evaluation and characterization of the PeanutMAGIC population.

Phenotypic evaluation for peanut pod and seed traits

The MAGIC Core and founders were planted at the USDA-ARS Belflower Farm in Tifton, GA, for two years, one plot in 2021 due to seed limitations and three replicate plots in 2022. The plot design consisted of two rows (1.2m long with 0.9m row space) separated by an alley of 1.5m, with a seeding rate of six seeds per 0.3m. Standard agronomic practices for growing peanut in Georgia were followed. After harvesting and drying to less than 10% moisture, the data were collected on a single plot basis for: (1) seed coat color (SCC), rated based on United States Peanut Descriptors (Pittman, 1995) as 0 to 6: where 0 = no color, 1 = white, 2 = tan, 3 = pink, 4 = red, 5 = purple, and 6 = dark purple; (2) pod weight (PW):100 pods were picked randomly, cleaned, and weighted for each line; (3) seed weight (SW) of 100 pods: the 100 pods were shelled then all seeds were cleaned, counted and weighted; and (4) shelling percentage (SP) was calculated by dividing the SW from the PW then multiplying by 100. The characteristics of pod constriction (PC) and pod reticulation (PR) were visually rated based on the United States Peanut Descriptors (Pittman, 1995) as PC ranged from 0 to 4: where 0 = none, 1 = slight, 2 = moderate, 3 = deep, and 4 = very deep and PR ranged from 1 to 4: where 1 = smooth, 2 = slight, 3 = moderate, and 4 = rough. Five technical replications of five seeds were used for each measurement of SCC, PC, and PR. The phenotypic data was analyzed and visualized using R Statistical Software (v4.4.0; R Core Team, 2021) using base function ‘mean’ and ‘sd’ to calculate arithmetic means and standard deviations. The function ‘ggplot’ was used to visualize the genotypic and phenotypic data. Similarly, the raw data and arithmetic means were input into TASSEL 5.0 for analysis.

DNA extraction, sequencing, and genotyping

Total genomic DNA was extracted from young leaflets of the founders and MAGIC Core using the CTAB method (Doyle and Doyle, 1987). The extracted DNA was checked for quality and quantity using Thermo Scientific E-Gel Power Snap Electrophoresis (Thermo Scientific, Wilmington, DE) and BMG LABTECH CLARIOstar (BMG LABTECH, Cary, NC). Samples without degradation and having a A_{260}/A_{280} ratio above 1.5 were standardized to 20ng/µL. The standardized DNA was sent to HudsonAlpha Institute for Biotechnology (Huntsville, AL) for library preparation and sequencing. The library was prepared using Twist 96-Plex Library Prep Kit (Twist Bioscience, South San Francisco, CA) followed by Illumina sequencing. The library consists of random oligos that act as primers to amplify random locations across the genome that were subsequently barcoded for multiplex sequencing. This generated low coverage (~1x) sequencing for SNP marker discovery. SNP calls were made using the Khufu pipeline (https://www.hudsonalpha.org/khufudata/) by aligning the sequencing reads to the peanut reference genome "Tifrunner" version 2 (https://www.peanutbase.org/). Khufu uses custom scripts to impute SNP calls to increase marker information across the population without a reference panel and is suitable for association analysis (Korani et al., 2021). A total of 138,151 SNPs were retained.

Genomic evaluation

A SNP density plot was constructed for the 20 Chr. using R/CMplot with 138,151 SNP markers (Yin et al., 2021). Markers are totaled in 1 Mb bins and plotted based on the physical position of the bin. Neighbor joining (NJ) trees were constructed using R/ape (Paradis and Schliep, 2018). First, the markers were made into a distance matrix using dist.gene, with default settings. The matrix was then converted into an NJ tree using nj, with default settings. The

parental lines were labeled and plotted using plot on R/ape. To extract the causality of the population structure, only SNPs on Chr. 9 and Chr. 19, hosting the introgression from wild peanut *A. cardenasii* (Chu et al., 2016; Clevenger et al., 2017), were removed then plotted using the same workflow. Both Chr. were selected based on the shared marker calls for the outgroup clade.

Minor allelic frequencies (MAF) were calculated using TASSEL 5 (Bradbury et al., 2007) then plotted using R/ggplot2 (Whickham, 2016). Linkage disequilibrium (LD) was calculated using TASSEL 5 through a sliding 100 marker window, then averaged in 500 Kb bins. The binned averages were plotted based on the physical position of the bin and respective Chr. using R/ggplot2 (Whickham, 2016). A sliding window approach allows for the calculation of local r^2 values with minimal influence from marker density variation across the Chr.

Genome-wide association study (GWAS)

GWAS was performed using 138,151 SNP markers and the phenotypic data for SCC, PW, SW, SP, PC, and PR by TASSEL 5.0 software. The SNP markers were first used to create a distance matrix for multidimensional scaling (MDS). MDS was calculated for two axes based on the elbow of the scree plot (Supplementary Figure 2.2). These 2 axes were used to control population structure during GWAS. The MDS, phenotype data, and genotyping data were joined by intersect join before performing a general linear model (GLM)-GWAS. For SCC, a significant threshold of $p < 1.0 \text{ X } 10^{-6}/N$, where N is the number of markers, was used. A Bonferroni significant threshold ($p < 0.01/N$) was used for PW, SW, and SP. For PC and PR, a significant threshold of $p < 1.0 \text{ X } 10^{-4}$ was used to identify low effect alleles, while minimizing false positives. The GWAS results were extracted from TASSEL 5.0 and plotted using R/CMplot.

Results

Development of PeanutMAGIC and genomic evaluation

Eight peanut cultivars and breeding lines were crossed in a simple funnel inter-cross breeding design as [A×B/C×D]/[E×F/G×H] through three generations of advanced inter-crossing and six selfing generations to produce 3,187 F_7 PeanutMAGIC RILs (Supplementary Figure 2.1). We made reciprocal crosses for each generation, particularly at 4-way and 8-way crosses, to produce more hybrid seeds. We produced 210 ABCD and 172 EFGH 4-way hybrid seeds (G_2) and 950 ABCDEFGH 8-way hybrid seeds (G_3) from 150 pairs of specific one-to-one parents. We paid close attention to the step from 8-way F_1 seeds to F_2 seeds, where each seed was identified by a unique coding number such as "39-9" (Figure 2.1). Heterosis or hybrid vigor was evidenced by larger F_1 plants compared to the border plants (Figure 2.1). A total of 3,575 F_2 seeds were harvested, then advanced to produce a total of 3,187 $F_{2:7}$ PeanutMAGIC RILs.

MAGIC Core has undergone low coverage Illumina sequencing along with the eight founder lines, yielding ~1x coverage for each line. Aligning the sequencing reads to the reference genome "Tifrunner" version 2 (http://www.peanutbase.org/www.peanutbase.org), we generated a total of 138,151 SNPs that were polymorphic among the eight founders and RILs. These SNPs were distributed across the whole genome, but the SNP density varied among and within the 20 Chr. (Figure 2.2A). Over 56% of the total SNPs were from Chr. 09 and Chr. 19, with 65,687 and 12,045 SNPs, respectively, as reported by Clevenger et al. (2017) that 92% of A09 was of *A. cardenasii* origin in the founder TifNV-H O/L (Holbrook et al., 2017) resulting in the high numbers of SNPs on these two homoeologous chromosomes. We found that the RILs and founders were not uniformly distributed from the center of the phylogenetic tree when

considering all SNPs, indicating the presence of a population structure (Figure 2.2B). The phylogenetic tree identified an outgroup of 52 outliers representing 16.8% of the RILs grouped with TifNV-H O/L (Figure 2.2B). The phylogenetic tree solely with Chr. 09 and Chr. 19 revealed the same outgroup and orientation with TifNV-H O/L as in Figure 2B. These results confirm the observed population structure originated from the introgressed genome segments of *A. cardenasii* (Supplementary Figure 2.3).

The distribution of the RILs was normalized and clustered relative to their parents when building the phylogenetic tree without SNPs from Chr. 09 and Chr. 19 (Figure 2.3A). We found that the RILs and the founders were not distributed at equal distances from the center of the phylogenetic trees. It was also interesting to note that parent Tifrunner was closer to FLA-07 when excluding SNPs from Chr. 09 and Chr. 19 (Figure 2.3A). In contrast, Tifrunner was closer to GA-13M than other founders when using all SNPs (Figure 2.2B). The minor allele frequencies possess a skewed distribution toward 0.12 or 12% throughout the population (Figure 2.3B). Excluding homoeologous Chr. 09 and Chr. 19, the allele frequency maintains a skewed distribution peaking at 0.1 or 10% (Figure 2.3C).

We identified high recombination events using r^2 values calculated in a sliding window of 100 markers, averaged in 500kb bins, and the r^2 values ranged from 0.001-0.406 (Figure 2.4A). Patterns of LD showed that the population had high recombination rates with low r^2 values (less than 0.2) along the chromosomes with several exceptions such as Chr. 01, Chr. 06, Chr. 09, Chr. 11, and Chr. 17. There are regions with a relatively lower recombination rate on Chr. 09 (Figure 2.4B), indicated by plateaued r^2 values. However, there were regions in Chr. 09 where multiple recombination points did occur as denoted by the valleys across the r^2 plot. These

locations align with the clades in the phylogenetic trees (Figure 2.4C). These data support that PeanutMAGIC is a diverse, highly recombinant synthetic population.

Pod and seed characteristics and variation

PeanutMAGIC population had several generations of intercrossing among the eight founders and produced greater genetic variation and novel allelic combinations resulting in diverse phenotypic traits as evidenced in pod and seed characteristics (Figure 2.5). To demonstrate the diversity, we selected two unique 8-way crosses and their derived $F_{2:3}$ families (Figure 2.5). One 8-way F_1 hybrid seed from crossing pair No. 50, coded as 50-10 (the 10^{th} seed), produced 12 $F_{2:3}$ families with contrasting morphologic pod shapes including PC, PR, and number of seeds per pod (Figure 2.5A). In contrast, crossing pair No. 43 produced 20 F_1 hybrid seeds and four $F_{2:3}$ families were selected from four different F_1 seeds (coded as 43-4, 43-5, 43-6, and 43-20) with different pod and seed characteristics (Figure 2.5B). Additionally, new diversity emerged, such as dark seed coat color and 3-seeded pod, characteristics that no founders possess (Figure 2.5).

PeanutMAGIC phenotypic variation was also evaluated by planting the MAGIC Core for two years and measuring the pod and seed traits including seed coat color (SCC), pod weight (PW), seed weight (SW), shelling percentage (SP), pod constriction (PC), and pod reticulation (PR). As expected, the broad genetic diversity produced wide phenotypic variation (Supplementary Table 2.1, Figure 2.6, Supplementary Figure 2.4). We observed continuous and transgressive phenotypic distributions for each trait except SCC that had a skewed distribution (Figure 2.6) (SCC Chi-square p-value < $2.2x10^{-16}$). Interestingly all founders have tan seed coat color (SCC = 2), but there were 78 lines (25.16%) with darker SCC (≥ 3). Furthermore,

PeanutMAGIC possesses lines that combine novel SCC and transgressive values of important seed traits (Supplementary Table 2.2).

Genome-wide association studies (GWAS) for peanut pod and seed traits

The genetic mapping power was assessed by GWAS of the MAGIC Core for each year (Supplementary Figure 2.5) and the averages for SCC, PW, SW, SP, PC, and PR (Figure 2.7, Figure 2.8, Figure 2.9, Supplementary Figure 2.5). Examining the numbers of significant markers from GWAS and phenotypic distributions, increasing complexity was observed among these traits from SCC to PR, respectively. SCC was identified as a qualitative trait (Figure 2.7, Supplementary Table 2.1, Supplementary Figure 2.4, Supplementary Figure 2.5). PW, SW, and SP were identified to be quantitative traits (Figure 2.8, Figure 2.9, Supplementary Table 2.1, Supplementary Figure 2.4, Supplementary Figure 2.5). PC and PR were identified as highly quantitative traits (Supplementary Table 2.1, Supplementary Figure 2.4, Supplementary Figure 2.5).

We identified significant SNPs associated with SCC within a linkage disequilibrium (LD) block of 10 Mb from 117.3 - 127.3 Mb (117,289,509 - 127,348,489) based on the "Tifrunner" version 2 reference genome on Chr. 03 with the $-\log_{10}(p)$ above the significant threshold of 12, using the two-year average data (Figure 2.7A). There was a minor difference in the identified regions between each year, 117.9 - 126.2 Mb (117,912,481 - 126,168,934) for year-1 and 117.9 - 127.3 Mb (117,912,481 - 127,348,489) for year-2, respectively (Supplementary Figure 2.5). We examined the GWAS signal pattern by enlarging Chr. 03 (Figure 2.7B) in comparison with the r^2 values (Figure 2.7C) and found that the associated region had an increase of LD (higher r^2) compared to neighboring regions (Figure 2.7C). In this LD block there were 430 annotated genes including three Myc-type bHLH transcription factors of *arahy.07ZIFT, arahy.4R4J32*, and

arahy.Y5CARX. The three SNP markers at 125.7 Mb (125,705,410), 126.1 Mb (126,095,135), and 126.2 Mb (126,168,934) had a *p*-value of 8.89 X 10^{-14}, 3.34 X 10^{-12}, and 9.32 X 10^{-17}, with a ecoefficiency of determination (R^2) of 0.17895, 0.15923, and 0.21495, respectively (Figure 2.7D). These three SNP markers are 138 Kb, 251 Kb, and 324 Kb, respectively, away from *arahy.07ZIFT* a Myc-type bHLH transcription factor (Figure 2.7D).

For both PW and SW, we identified significant SNPs associated within a 2.4 Mb region from 105.4 – 107.8 Mb (105,403,922 - 107,767,685) on Chr. 05 with the -$\log_{10}$ (*p*) above the significant threshold of 7 (Figure 2.8, Supplementary Figure 2.5). PW and SW are strongly correlated with a Pearson correlation of 0.966. We examined the GWAS signal patterns by enlarging the Chr. and discovered that there were three separate regions (Figure 2.8B, 2.8D, as color coded). Both PW and SW shared two regions, 105.4 - 105.7 Mb (105,403,922 - 105,697,089) and 106.9 - 107.8 Mb (106,891,348 - 107,767,685) (Figure 2.8), and the middle region showed minor variation as 106.2 - 106.3 Mb (106,210,675 - 106,301,816) for PW and 106.2 - 106.5 Mb (106,210,675 - 106,502,469) for SW (200 Kb larger) (Figure 2.8B, 2.8D, coded as yellow color). For the PW average, we also identified a significant region on Chr. 07 from 1.3 - 1.9 Mb (1,282,902 - 1,904,787) above the Bonferroni threshold (Figure 2.8A). In the three co-associated regions, there were 16 annotated genes from 105.4 - 105.7 Mb, 19 annotated genes from 106.2 - 106.5 Mb, and 36 annotated genes from 106.9 - 107.8 Mb on Chr. 05. In total, there were 71 genes within the significant regions on Chr. 05 for PW and SW. There was a potential candidate gene called spermidine synthase associated with PW and SW (*arahy.DY91G5*, 107,419,666 - 107,427,823) (Gangurde et al., 2019).

Using the SP average data, we identified significantly associated SNPs within a 30 Kb region (27,002,857 - 27,038,220) on Chr. 05 (Figure 2.9A), where there had no annotated genes.

The nearest gene was *arahy.G26LFA*, a CheY-like response regulator 3. For the yearly SP, there was no significant associated SNPs with the year-1 data, but the year-2 SP identified a 10 Mb region from 18,494,757 - 28,460,594 on Chr. 05 (Supplementary Figure 2.5), containing 144 annotated genes. The markers within the 30 Kb region come from the founder NC94022, a low SP breeding line. It was interesting that the associated flanking and adjacent regions exhibited strong GWAS signals (Figure 2.9B), indicating that this location had high LD for high shelling percentage lines, but the MAGIC Core population had a low LD (smaller r^2 value) within this area (Figure 2.9C).

For peanut pod shape, we measured two highly quantitative traits, PC and PR, for two years. GWAS analysis showed that there were more significant SNP locations associated with PC and PR throughout the genome. There were only four locations consistently associated with both years and the average data for PC and PR, respectively (Supplementary Figure 2.5). The consistent four locations for PC were: one region on Chr. 06 from 52,744 - 61,427 and three regions colocalized on Chr. 18 from 56,805,456 - 56,894,108, 76,668,672 - 76,669,371, and 109,119,479 - 109,119,520, all without annotated genes Supplementary (Figure 2.5). The consistent four locations for PR were on different Chr., which were: Chr. 02 from 90,047,749 - 90,194,264; Chr.05 from 9,537,177 - 9,626,562; Chr. 08 from 18,088,246 - 18,088,909; and Chr.18 from 6,923,065 - 7,267,864 (Supplementary Figure 2.5). There were 35 annotated genes within these regions for PR.

Discussion

PeanutMAGIC construction and genomic qualities

Multiparent advanced generation inter-cross (MAGIC) populations leverage multiple genetic sources and meiotic recombination to rearrange distinct alleles from all founders equally. Therefore, MAGIC populations could produce novel allele combinations resulting in unique recombinants to serve as a new genetic resource for high-resolution mapping and for plant breeding. For this purpose, we reported a peanut MAGIC population, PeanutMAGIC, with 3,187 RILs developed from eight diverse peanut cultivars and breeding lines. In this study, we used a subset of 310 RILs, called MAGIC Core, and conducted genotypic and phenotypic studies to evaluate and characterize PeanutMAGIC. The analysis of the MAGIC Core revealed that PeanutMAGIC possesses a balanced and uniformly differentiated mosaic of genome blocks, each originating from one of the eight founders. This provided an array of unique recombination events, which facilitated high-definition mapping of traits that extend from qualitative to highly quantitative and offers substantial implications for breeding purposes.

It is important for multiparent populations to remain balanced during construction to yield an equal distribution of genotypes and phenotypes. Therefore, during PeanutMAGIC development, the 8-way crossing posed the greatest challenge, being the most critical, difficult, and time-consuming phase. For preparation of the 8-way crossing, we produced more both 4-way hybrids which was necessary for the 8-way crossing to maximize both the size of the population and potential combinations of alleles. We also tracked each 8-way hybrid seed with two numbers, the 8-way pair number and the resulting F_1 hybrid seed in numerical order (Figure 2.1) till the $F_{2:7}$ generation to ensure there were an equal number of $F_{2:7}$ families in the final PeanutMAGIC population. These breeding techniques allowed for high recombination rates of chromosomes while maintaining equal representation of the eight founders within the population, yielding a diverse, synthetic, balanced population with known pedigrees. The concerted efforts

have culminated in the production of broad phenotypic variation in both pod and seed traits, exhibiting significant transgressive segregation within PeanutMAGIC population. This variation showcases high phenotypic diversity and underscores the potential of PeanutMAGIC as a resource for selection of breeding lines and cultivars.

The random selection method of the MAGIC Core allowed for the identification of general features within PeanutMAGIC, without creating artificial population structure. The MAGIC Core subset can be utilized for specific research applications to accumulate phenotypes, while leveraging consistent genotyping data over time. The presence of high-density SNPs on the homoeologous Chr. 09 and Chr.19 is attributed to the introgressed genome segments from the founder TifNV-H O/L (Holbrook et al., 2017) originating from the diploid peanut *A. cardenasii,* which is known to confer resistance to root-knot nematode (RKN). However, this introgression has exhibited very limited recombination in bi-parent breeding populations (Chu et al., 2016; Clevenger et al., 2017). The identification of TifNV-H O/L within the excluded cluster in Figure 2B confirms that the skewed population structure stems from the known *A. cardenasii* introgression (Holbrook et al., 2017). The different clades within the TifNV-H O/L outgroup represent novel recombinants of the introgression, which may aid in root-knot nematode resistance loci identification at higher resolutions than previously achieved (Chu et al., 2016; Clevenger et al., 2017). The increased polymorphism on Chr. 19 is associated with RILs that possess the introgression on Chr. 09, suggesting that homoeologous recombination of the introgression may occur. Smaller fragments of the introgression undergoing translocation may explain the slight divergence of the clade with the founder TifNV-H O/L.

The genetic distances and allele frequencies excluding Chr. 09 and Chr. 19 markers demonstrate the balance and the diversity within PeanutMAGIC without artificial inflation from

the wild diploid introgression. An even distribution of the founders in the NJ tree exemplifies the balance between each founder. Tifrunner and FLA-07 in close proximity underline the limited genetic diversity of peanut cultivars and stringent requirements for both quality and disease resistance, suggesting broadening peanut germplasm is needed (Brown et al., 2021). Alleles from homoeologous Chr. 09 and Chr. 19 harbor increased diversity from introgressed genome segments of *A. cardenasii*. Excluding Chr. 09 and Chr. 19, the minor allele frequency maintains a skewed distribution peaking at 0.1 or 10%, suggesting that PeanutMAGIC was a diverse synthetic population near theoretic proportions (12.5%) without inflation from introgressed alleles. The low r^2 values across the 20 Chr. show all regions of the genome are recombining, including introgressed segments on Chr. 09 and pericentromeric regions. The recombination of traditionally non-recombining regions at a contribution level near theoretical proportions demonstrates the high recombination and balance this population possesses. Therefore, PeanutMAGIC has a balanced and evenly differentiated mosaic of chromosomal segments from the eight founders, making it an excellent resource for genomic studies and potential for breeding.

PeanutMAGIC phenotypic variation and possible as breeding resource

The MAGIC Core exhibits broad and transgressive segregation for all traits examined in this study. This is the physical evidence of high genomic recombination and can be applied for genetic gain. Novel phenotypes from the founders have been identified within PeanutMAGIC such as SCC and 3-seeded pods. This spontaneous phenotypic change could be the product of genomic rearrangement, through homoeologous recombination or transposable element activity such as peanut flower color changes reported by Bertioli et al. (2019). Within the population, SCC spontaneously changed from tan into darker pigments. This is a change in regulation of

anthocyanin production, the main cause of pigmentation in peanut seed coats (Sarnoski et al., 2012; Zhao et al. 2017). The identification of new phenotypes within this population exemplifies the novel genetic variation within PeanutMAGIC for the improvement of peanut genetic resources.

The normal, transgressive segregation for pod and seed traits demonstrates the broad genetic variation within PeanutMAGIC. Interestingly, the phenotypes of the parental lines exhibit comparable scores for many traits examined in this study. The transgressive segregation observed in PeanutMAGIC phenotypes further emphasizes the prevalence of epistatic effects within this population. There are lines with transgressive and novel traits throughout PeanutMAGIC, such as MG505 with purple SCC scored as 5, high SP (77.4%, 81st percentile), very deep constriction (scored as 4), and moderate reticulation (scored as 3). Furthermore, there are other lines that express a combination of novel and commercially desired traits within PeanutMAGIC, including MG1206 with red SCC (scored as 4), high SP (77.1%), no constriction, and smooth reticulation. These characteristics are noteworthy for both genomic studies and for increased diversity of the limited gene pool of cultivated peanuts (Bertioli et al., 2019).

Enhancing genetic gain with refined mapping

To experimentally assess the mapping power of the PeanutMAGIC, pod and seed traits were used for GWAS analysis. The increased complexity of these traits facilitated the assessment of PeanutMAGIC genomic mapping capabilities for various traits with different inheritance patterns. A GLM approach was used based on the normal distribution of phenotypes and limited population structure, complex model approaches unwarranted. The population structure from the *A. cardenasii* introgression on Chr. 09 was corrected through the MDS fixed effect variable in

the linear model. Over correcting the population structure and random effects reduces the signal from markers, however, underestimating these effects creates false positives. Thus, careful consideration was taken to ensure proper parameters were used in data analysis.

In this study, we identified a novel SCC trait that is not possessed by any founders. GWAS analysis identified a significant region from 117.3-127.3 Mb on Chr. 03, harboring 430 annotated genes. The non-normal distribution of the phenotypes inflated the significance of the associated markers and flanking regions with measurable LD. Furthermore, markers in the pericentromeric region display a similar pattern in the GWAS signal as the r^2 plot, showing the sensitivity of GWAS to LD. The increased LD in the identified region indicates that recombination was impeded in this area. This might suggest a structural variation is linked or the causality of this phenotype, formed during population development. Reduced recombination rates caused by structural variants have been described previously within the rice pan-genome inversion index (Zhou et al., 2023). In the region from 125.7 – 126.1 Mb (Figure 2.7D), there were three significant SNPs at 126,168,934 with the highest value of –log10 (*p)*, 126,095,135, and 125,705,410, respectively, in the flanking regions linked with one candidate gene, a Myc-type bHLH transcription factor (*arahy.07ZIFT*) from 125,840,913 to 125,843,766 bp (Figure 2.7D). Myc-type bHLH transcription factors have been reported to regulate anthocyanin production in other plant seeds such as *Arabidopsis*, rice, and morning glory (Nesi et al., 2000; Sweeney et al., 2006; Park et al., 2007). For peanut SCC, several studies report the identification of one major QTL on A03 (Chr. 03) (Shirasawa et al., 2012; Zhuang et al., 2019; Chen et al., 2021). Chen et al. (2021) reported the same gene *arahy.07ZIFT* as the putative candidate gene as *Arachis hypogaea Red Testa 1* (*AhRt1*) through sequence analysis and functional annotation of a biparental population. Parallel identification of a functional gene through spontaneous

phenotypic change within PeanutMAGIC proves the suitability of this population for fine-mapping and candidate gene discovery.

PW and SW are polygenic traits controlled by several genes in other species (Han et al., 2012; Liu et al., 2015). In this study, GWAS analyses identified the same region on Chr. 05 broadly from 105.4-107.8 Mb, for PW and SW in both years and the average. Within this broad region is a cluster of three significantly associated regions allowing for further fine-mapping of this QTL. The loci associated with PW and SW have been reported in several previous studies (Luo et al., 2017, Chu et al., 2019b, and Gangurde et al., 2020). These studies identified a strong locus on A05 (Chr. 05), in a parallel location using the founder genotypes of PeanutMAGIC (Chu et al., 2019b; Gangurde et al., 2019). There was one reported candidate gene termed spermidine synthase associated with PW and SW, also identified in this study (*arahy.DY91G5*, 107,419,666 - 107,427,823) (Gangurde et al., 2019). The transition from a broad region to three pin-point regions underscores the enhancement in fine-mapping and potential for candidate gene identification.

Shelling percentage (SP) in peanut is an important agronomic trait second only to the yield. For SP, we identified a significant associated region from 27.0-27.03 Mb on Chr. 05 for both years and the average. The GWAS signal of distal neighboring regions are elevated, suggesting that the high SP haplotype is recalcitrant to recombine. Interestingly, the LD values for the associated region among the population is low, indicating the region is selectively recombining. This suggests the causal allele is only complementary to specific haplotypes within the population. A cause for non-uniform recombination may be from structural variants that do not always have compatible haplotypes. Structural variants are strong contenders for causing trait variation, particularly within species less divergent from their progenitors (Li et al., 2023). Luo

et al. (2018) reported associated SNPs on chromosome A09 (Chr. 09) and B02 (Chr. 12) using QTL-seq. This was likely due to different parental lines used. We found that the significant SNP locations are from the founder NC94022, suggesting that this breeding line has distinct adverse alleles for SP. This information is crucial for understanding the interactions of beneficial and detrimental alleles within breeding lines for efficient genetic gain in breeding programs.

PC and PR are highly quantitative traits as evidenced by the phenotypic statistics and GWAS analyses. There were two reports on PC and PR mapping using biparental populations. Patil et al. (2018) only identified B07 (Chr. 17) for PC. Mondal and Badigannavar (2019) identified B07 and B08 (Chr. 18) for PC and B08 for PR. In this study we found 4 regions for PC, 3 on Chr. 18 (B08) and for PR we also identified 4 regions, 1 on Chr. 18. This study agreed with previous loci and identified novel locations for PC and PR. Moreover, the support of previous loci and identification of new loci for highly quantitative traits demonstrates the utility of PeanutMAGIC for understanding complex and understudied traits.

GWAS analyses for SCC, PW, SW, SP, PC, and PR have found consistently associated SNPs within genome regions that agreed with previous studies and identified novel loci. However, the increased resolution mapping of these regions shows the ability of PeanutMAGIC to map a spectrum of qualitative to highly quantitative traits effectively. Therefore, PeanutMAGIC is a valuable resource for further trait mapping of complex traits to understand and improve genetic gain in cultivated peanut breeding.

In summary, as the resolution of genetic mapping in a relatively small sized bi-parental population is usually insufficient in identifying the causal variants underlying QTLs, the development of muti-parent population like PeanutMAGIC has been used to address these concerns. MAGIC populations improve the precision of QTL mapping over bi-parental

populations by incorporating increased diversity and opportunities to reduce linkage disequilibrium among variants. The PeanutMAGIC will serve as a unique genetic and genomic resource to characterize the genetics of complex traits and potentially identify superior alleles for crop improvement in breeding. Therefore, this study reports the assessment of the first multiple-parent advanced generation inter-cross population in peanut, the PeanutMAGIC, and its unique properties in genetic mapping of complex traits. PeanutMAGIC has broadened genetic diversity through a balanced and differentiated mosaic of eight founder haplotypes and serves as a novel genetic resource for breeding applications. This study also demonstrated that PeanutMAGIC can achieve high resolution mapping for several known and understudied traits. PeanutMAGIC and MAGIC Core are available to researchers for research and breeding use upon request.

References

Agarwal, G., Clevenger, J., Pandey, M.K., Wang, H., Shasidhar, Y., Chu, Y., Fountain, J.C., Choudhary, D., Culbreath, A.K., Liu, X., Huang, G., Wang, X., Deshmukh, R., Holbrook, C.C., Bertioli, D.J., Ozias-Akins, P., Jackson, S.A., Varshney, R.K. & Guo, B.Z. (2018). High-density genetic map using whole-genome re-sequencing for fine mapping and candidate gene discovery for disease resistance in peanut. *Plant Biotechnology Journal*, 16(11), 1954-1967,16:1954-1967. https://doi.org/10.1111/pbi.12930

Agarwal, G., Clevenger, J., Kale, S.M., Wang, H., Pandey, M.K., Choudhary, D., Yuan, M., Wang, X., Culbreath, A.K., Holbrook, C.C., Liu, X., Varshney, R.K. & Guo, B.Z. (2019). A recombination bin-map identified a major QTL for resistance to Tomato Spotted Wilt Virus in peanut (*Arachis hypogaea*). *Scientific Reports. 9:18246* | https://doi.org/10.1038/s41598-019-54747-1

Arrones, A., Vilanova, S., Plazas, M., Mangino, G., Pascual, L., Díez, M. J., Prohens, J., & Gramazio, P. (2020). The dawn of the age of multi-parent magic populations in plant breeding: Novel powerful next-generation resources for genetic analysis and selection of recombinant elite material. *Biology*, *9*(8), 229. https://doi.org/10.3390/biology9080229

Bertioli, D. J., Cannon, S. B., Froenicke, L., Huang, G., Farmer, A. D., Cannon, E. K. S., Liu, X., Gao, D., Clevenger, J., Dash, S., Ren, L., Moretzsohn, M. C., Shirasawa, K., Huang, W., Vidigal, B., Abernathy, B., Chu, Y., Niederhuth, C. E., Umale, P., Araujo, A. C. G., Kozik, A., Kim, K. D., Burow, M. D., Varshney, R. K., Wang, X., Zhang, X., Barkley, N., Guimaraes, P. M., Isobe, S., Guo, B., Liao, B., Stalker, H. T., Schmitz, R. J., Scheffler, B. E., Leal-Bertioli, S. C. M., Xun, X., Jackson, S., A Michelmore, R., & Ozias-Akins, P. (2016). The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nature Genetics*, *48*, 438– 446. https://doi.org/10.1038/ng.3517

Bertioli, D. J., Jenkins, J., Clevenger, J., Dudchenko, O., Gao, D., Seijo, G., Leal-Bertioli S. C. M., Ren, L., Farmer, A. D., Pandey, M. K., Samoluk, S. S., Abernathy, B., Agarwal, G., Ballen-Taborda, C., Cameron, C., Campbell, J., Chavarro, C. Chitikineni, A., Chu, Y., Dash, S., Baidouri, M. E., Guo, B., Huang, W., Kim, K. D., Korani, W., Lanciano, S., Lui, C. G., Mirouze, M., Moretzsohn, M. C., Pham, M., Shin, J. H., Shirasawa, K., Sinharoy, S., Sreedasyam, A., Weeks, N. T., Zhang, X., Zheng, Z., Sun, Z., Froenicke, L., Aiden, E. L., Michelmore, R., Varshney, R. K., Holbrook, C. C., Cannon, E. K. S., Scheffler, B. E., Grimwood, J., Ozias-Akins, P., Cannon S. B., Jackson, S. A., & Schmutz, J. (2019). The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nature Genetics 51*, 877–884. https://doi.org/10.1038/s41588-019-0405-z

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

Branch, W. D. (2007). Registration of 'Georgia-06G' Peanut. *Journal of Plant Registrations 1(2)*: 120. https://doi.org/10.3198/jpr2006.12.0812crc

Branch, W. D. (2014). Registration of 'Georgia-13M' Peanut. *Journal of Plant Registrations 8(3)*: 253–256. https://doi.org/10.3198/jpr2013.11.0071crc

Brown N., Branch W. D., Johnson M., & Wallace J. (2021). Genetic diversity assessment of Georgia peanut cultivars developed during ninety years of breeding. *Plant Genome 14(3)*: e20141. https://doi.org/10.1002/tpg2.20141

Chavarro, C., Chu, Y., Holbrook, C., Isleib, T., Bertioli, D., Hovav, R., Butts, C., Marshall, L., Sorensen, R., Jackson, S.A., & Ozias-Akins, P. (2020). Pod and seed trait QTL identification to assist breeding for peanut market preferences. *G3: Genes, Genomes, Genetics. 10*, 2297-2315. https://doi.org/10.1101/738914

Chen, H., X. Chen, R. Xu, W. Liu, N. Liu, L. Huang, H. Luo, D. Huai, X. Lan, Y. Zhang, R. Hu, J. Chen, Z. Tang, G. Lin, & H. Jiang. (2021). Fine-mapping and gene candidate analysis for Ahrt1, a major dominant locus responsible for Testa color in cultivated peanut. *Theoretical and Applied Genetics, 134,* 3721–3730. https://doi.org/10.1007/s00122-021-03924-w

Chen, X., H. Li, M.K. Pandey, Q. Yang, X. Wang, V. Garg, H. Li, X. Chi, D. Doddamani, Y. Hong, H. Upadhyaya, H. Guo, A.W. Khan, F. Zhu, X. Zhang, L. Pan, G.J. Pierce, G. Zhou, K.A. Krishnamohan, M. Chen, N. Zhong, G. Agarwal, S. Li, A. Chitikineni, G.-Q. Zhang, S. Sharma, N. Chen, H. Liu, P. Janila, S. Li, M. Wang, T. Wang, J. Sun, X. Li, C. Li, M. Wang,

L. Yu, S. Wen, S. Singh, Z. Yang, J. Zhao, C. Zhang, Y. Yu, J. Bi, X. Zhang, Z.-J. Liu, A.H. Paterson, S. Wang, X. Liang, R.K. Varshney, & S. Yu. (2016). Draft genome of the peanut a-genome progenitor *Arachis duranensis* provides insights into geocarpy, oil biosynthesis, and allergens. *Proceedings of the National Academy of Sciences 113(24)*: 6785–6790. https://doi.org/10.1073/pnas.1600899113

Chen, X., Q. Lu, H. Liu, J. Zhang, Y. Hong, H. Lan, H. Li, J. Wang, H. Liu, S. Li, M.K. Pandey, Z. Zhang, G. Zhou, J. Yu, G. Zhang, J. Yuan, X. Li, S. Wen, F. Meng, S. Yu, X. Wang, K.H.M. Siddique, Z.-J. Liu, A.H. Paterson, R.K. Varshney, & X. Liang. (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

Chu, Y., Chee, P., Culbreath, A. K., Isleib, T. B., Holbrook Jr, C. C., & Ozias-Akins, P. (2019a). Major QTLs for resistance to early and late leaf spot diseases are identified on chromosomes 3 and 5 in peanut (*Arachis hypogaea*). *Frontiers in Plant Science*. *10*:1-13. https://doi.org/10.3389/fpls.2019.00883

Chu, Y., Chee, P., Isleib, T. G., Holbrook, C. C., & Ozias-Akins, P. (2019b). Major seed size QTL on chromosome A05 of peanut (*Arachis hypogaea*) is conserved in the US mini core germplasm collection. *Molecular Breeding, 40(1)*. https://doi.org/10.1007/s11032-019-1082-4

Chu, Y., Gill, R., Clevenger, J. P., 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

Clevenger, J., Chu, Y., Arrais Guimaraes, L., Maia, T., Bertioli, D., Leal-Bertioli, S., Timper, P., Holbrook, C.C., & Ozias-Akins, P. (2017). Gene expression profiling describes the genetic regulation of *Meloidogyne arenaria* resistance in *Arachis hypogaea* and reveals a candidate gene for resistance. *Scientific Reports 7*, 1317 https://doi.org/10.1038/s41598-017-00971-6

Clevenger, J. P., & P. Ozias-Akins. (2015a). SWEEP: A Tool for Filtering High-Quality SNPs in Polyploid Crops. *G3: Genes, Genomes, Genetics 5(9)*: 1797–1803. https://doi.org/10.1534/g3.115.019703

Clevenger, J., C. Chavarro, S.A. Pearl, P. Ozias-Akins, & S.A. Jackson. (2015b). Single nucleotide polymorphism identification in polyploids: A review, example, and recommendations. *Molecular Plant 8(6)*: 831–846.

Culbreath, A. K, Gorbet, D. W., Martinez-Ochoa, N., Holbrook, C. C., Todd, J. W., Isleib, T. G., & Tillman, B. (2005). High levels of field resistance to tomato spotted wilt virus in peanut breeding lines derived from *hypogaea* and *hirsuta* botanical varieties. *Peanut Science 32*:20–24. https://doi.org/10.3146/0095-3679.

Dell'Acqua, M., Gatti, D. M., Pea, G., Cattonaro, F., Coppens, F., Magris, G., Hlaing, A. L., Aung, H. H., Nelissen, H., Baute, J., Frascaroli, E., Churchill, G. A., Inzé, D., Morgante, M., & Pè, M. E. (2015). Genetic properties of the magic maize population: A new platform for high definition QTL mapping in *Zea mays*. *Genome Biology*, *16*(1). https://doi.org/10.1186/s13059-015-0716-z

Doyle, J.J. & Doyle, J.L. (1987). A rapid DNA isolation for small quantities of fresh leaf tissue. *Phytochemistry Bulletin 19:*11–15

Gangurde, S. S., Wang, H., Yaduru, S., Pandey, M. K., Fountain, J. C., Chu, Y., Isleib, T., Holbrook, C. C., Xavier, A., Culbreath, A. K., Ozias-Akins, P., Varshney, R. K., & Guo, B. (2019). Nested-Association mapping (NAM)-based genetic dissection uncovers candidate genes for seed and pod weights in peanut (*Arachis hypogaea). Plant Biotechnology Journal, 18*(6), 1457–1471. https://doi.org/10.1111/pbi.13311

Gangurde, S.S., E. Thompson, S. Yaduru, H. Wang, J.C. Fountain, Y. Chu, P. Ozias-Akins, T.G. Isleib, C.C. Holbrook, B. Dutta, A.K. Culbreath, M.K. Pandey, & B. Guo. (2024). Linkage-mapping and genome-wide association study identified two peanut late leaf spot resistance loci, *PLLSR*-1 and *PLLSR*-2, using a nested association mapping. *Phytopathology 114(6)*, 1346-1355. https://doi.org/10.1094/PHYTO-04-23-0143-R

Gorbet, D. W., & B. L. Tillman (2009). Registration of 'Florida-07' Peanut. *Journal of Plant Registrations 3(1)*: 14–18. https://doi.org/10.3198/jpr2008.05.0276crc

Gorbet, D. W., & Knauft, D. A. (2000). Registration of 'SunOleic 97R' peanut. *Crop Science 40*:1190–1191. https://doi.org/10.2135/cropsci2000.0032rcv

Guo, B. Z., **C**hen, C. Y., Chu, Y., Holbrook, C. C., Ozias-Akins, P. & Stalker, H. T. (2012). Advances in genetics and genomics for sustainable peanut production. In: Benkeblia, N., editor. Sustainable Agriculture and New Biotechnologies, Boca Raton, FL: CRC Press. p. 341-367. https://doi.org/10.1201/b10977-20

Guo, B.Z., Khera, P., Wang, H., Peng, Z., Sudini, H., Wang, X., Osiru, M., Chen, J., Vadez, V., Yuan, M., Wang, C.T., Zhang, X., Waliyar, F., Wang, J. & Varshney, R.K. (2016). Annotation of trait loci on integrated genetic maps of *Arachis* species. In: Stalker, H.T., Wilson, R.F., editors. Peanuts: Genetics, Processing, and Utilization. Academic Press and the

American Oil Chemists' Society (AOCS) Press, pp. 163-207. https://doi.org/10.1016/b978-1-63067-038-2.00006-x

Han, Y., Li, D., Zhu, D., Li, H., Li, X., Teng, W., & Li, W. (2012). QTL analysis of soybean seed weight across multi-genetic backgrounds and environments. *Theoretical and Applied Genetics, 125(4),* 671–683. https://doi.org/10.1007/s00122-012-1859-x

Holbrook, C. C., Anderson, W. F., & Pittman, R. N. (1993). Selection of a core collection from the U. S. germplasm collection of peanut. *Crop Science*, *33,* 859–861.

Holbrook, C. C., & Culbreath, A. K. (2007). Registration of 'Tifrunner' peanut. *Journal of Plant Registrations, 1(2),* 124–124.

Holbrook, C. C., & Dong, W. (2005). Development and evaluation of a mini core collection for the U.S. peanut germplasm collection. *Crop Science*, *45*,1540–1544. https://

Holbrook, C. C., Isleib, T. G., Ozias-Akins, P., Chu, Y., Knapp, S. J., Tillman, B., Guo, B., Gill, R., & Burow, M. D. (2013). Development and phenotyping of recombinant inbred line (RIL) populations for peanut (*Arachis hypogaea*). *Peanut Science*, *40,* 89-94. https://doi.org/10.3146/PS13-5.1

Holbrook, C. C., Ozias-Akins, P., Chu, Y., Culbreath, A. K., Kvien, C. K., & Brenneman, T. B. (2017). Registration of 'TifNV-high O/L' peanut. *Journal of Plant Registrations*, *11*(3), 228–230. https://doi.org/10.3198/jpr2016.10.0059crc

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

Husted, L. (1936). Cytological studies on the peanut, Arachis. II. Chromosome number, morphology and behavior, and their application to the problem of the origin of the cultivated forms. *Cytological, 7,* 396–23. https://doi.org/10.1508/cytologia.7.396

Isleib, T. G., Holbrook, C. C., & Gorbet, D. W. (2001). Use of plant introductions in peanut cultivar development. *Peanut Science, 28,* 96–113. https://doi.org/10.3146/i0095-3679-28-2-11

Khera, P., M.K. Pandey, H. Wang, S. Feng, L. Qiao, A.K. Culbreath, S. Kale, J. Wang, C.C. Holbrook, W. Zhuang, R.K. Varshney, & B. Guo. (2016). Mapping quantitative trait loci of resistance to tomato spotted wilt virus and leaf spots in a recombinant inbred line population of peanut (*Arachis hypogaea* L.) from SunOleic 97R and NC94022. *PLOS ONE 11(7).* https://doi.org/10.1371/journal.pone.0158452

Korani, W., O'Connor, D., Chu, Y., Chavarro, C., Ballen, C., Guo, B., Ozias-Akins, P., Wright, G., & Clevenger, J. (2021). De Novo QTL-seq identifies loci linked to blanchability in peanut (*Arachis hypogaea*) and refines previously identified QTL with low coverage sequence. *Agronomy, 11*(11), 2021. https://doi.org/10.3390/agronomy11112201

Li, N., He, Q., Wang, J., Wang, B., Zhao, J., Huang, S., Yang, T., Tang, Y., Yang, S., Aisimutuola, P., Xu, R., Hu, J., Jia, C., Ma, K., Li, Z., Jiang, F., Gao, J., Lan, H., Zhou, Y., Zhang, X., Huang, S., Fei, Z., Wang, H., Li, H., & Yu, Q. (2023). Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. *Nature Genetics, 55*(5), 852–860. https://doi.org/10.1038/s41588-023-01340-y

Liang, X., Holbrook, C. C., Lynch, R. E., & Guo, B. Z. (2005) Beta-1,3-glucanase activity in peanut seed (*Arachis hypogaea*) is induced by inoculation with *Aspergillus flavus* and

copurifies with a conglutin-like protein**.** *Phytopathology 95,* 506-511 https://doi.org/10.1094/PHYTO-95-0506

Liu, J., Hua, W., Hu, Z., Yang, H., Zhang, L., Li, R., Deng, L., Sun, X., Wang, X., & Wang, H. (2015). Natural variation in *ARF18* gene simultaneously affects seed weight and silique length in polyploid rapeseed. *Proceedings of the National Academy of Sciences, 112(37).* https://doi.org/10.1073/pnas.1502160112

Luo, H., Guo, J., Ren, X., Chen, W., Huang, L., Zhou, X., Chen, Y., Liu, N., Xiong, F., Lei, Y., Liao, B., & Jiang, H. (2017). Chromosomes A07 and A05 associated with stable and major QTLs for pod weight and size in cultivated peanut (*Arachis hypogaea* L.). *Theoretical and Applied Genetics*, *131*(2), 267–282. https://doi.org/10.1007/s00122-017-3000-7

Luo, H., Pandey, M. K., Khan, A. W., Guo, J., Wu, B., Cai, Y., Huang, L., Zhou, X., Chen, Y., Chen, W., Liu, N., Lei, Y., Liao, B., Varshney, R. K., & Jiang, H. (2018). Discovery of genomic regions and candidate genes controlling shelling percentage using QTL-seq approach in cultivated peanut (*Arachis hypogaea* L.). *Plant Biotechnology Journal*, 17(7), 1248–1260. https://doi.org/10.1111/pbi.13050

Luo, M., P. Dang, B.Z. Guo, G. He, C.C. Holbrook, M.G. Bausher, & R.D. Lee. (2005). Generation of expressed sequence tags (ESTs) for gene discovery and marker development in cultivated peanut. *Crop Science 45(1)*: 346–353. https://doi.org/10.2135/cropsci2005.0346

Luo, Z., Cui, R., Chavarro, C., Tseng, Y., Zhou, H., Peng, Z., Chu, Y., Yang, X., Lopez, Y., Tillman, B., Dufault, N., Brenneman, T., Isleib, T.G., Holbrook Jr, C.C., Ozias-Akins, P., & Wang, J. (2020). Mapping quantitative trait loci (QTLs) and estimating the epistasis controlling stem rot resistance in cultivated peanut (*Arachis hypogaea*). *Journal of*

Theoretical and Applied Genetics. *133*, 1201-1212. https://doi.org/10.1007/s00122-020-03542-y

Mondal, S., & Badigannavar, A. M. (2019). Identification of major consensus QTLs for seed size and minor QTLs for pod traits in cultivated groundnut (*Arachis hypogaea* L.). *3 Biotech*, *9*(9). https://doi.org/10.1007/s13205-019-1881-7

Nesi, N., I. Debeaujon, C. Jond, G. Pelletier, M. Caboche, & L. Lepiniec. (2000). The TT8 gene encodes a basic helix-loop-helix domain protein required for expression of DFR and ban genes in Arabidopsis siliques. *The Plant Cell 12(10)*: 1863. https://doi.org/10.2307/3871198

Pandey, M.K., H. Wang, P. Khera, M.K. Vishwakarma, S.M. Kale, A.K. Culbreath, C.C. Holbrook, X. Wang, R.K. Varshney, & B. Guo. (2017). Genetic dissection of novel QTLs for resistance to leaf spots and tomato spotted wilt virus in peanut (*Arachis hypogaea* L.). *Frontiers in Plant Science 8*. https://doi.org/10.3389/fpls.2017.00025

Paradis, E., & Schliep, K. (2018). APE 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Park, K., N. Ishikawa, Y. Morita, J. Choi, A. Hoshino, & S. Iida. (2007). A bHLH regulatory gene in the common morning glory, *Ipomoea purpurea*, controls anthocyanin biosynthesis in flowers, proanthocyanidin and phytomelanin pigmentation in seeds, and seed Trichome Formation. *The Plant Journal*, *49*, 641–654. https://doi.org/10.1111/j.1365-313X.2006.02988.x

Patil, A. S., Popovsky, S., Levy, Y., Chu, Y., Clevenger, J., Ozias-Akins, P., & Hovav, R. (2018). Genetic insight and mapping of the pod constriction trait in Virginia-type peanut. *BMC Genetics, 19*(1). https://doi.org/10.1186/s12863-018-0674-z

Pittman, R. (1995, January 1). *United States Peanut Descriptors: Pittman, R.N: Free download, Borrow, and streaming.* Internet Archive. https://archive.org/details/IND20479053/page/n1/mode/2up

Qin, H., Feng, S., Chen, C., Guo, Y., Knapp, S., Culbreath, A., He, G., Wang, M.L., Zhang, X., Holbrook, C.C., Ozias-Akins, P., & Guo, B. (2012). An integrated genetic linkage map of cultivated peanut (*Arachis hypogaea* L.) constructed from two RIL populations. *Theoretical and Applied Genetics, 124(4)*: 653–664. https://doi.org/10.1007/s00122-011-1737-y

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical computing, Vienna, Austria. https://www.R-project.org/.

Roy, J., Soler-Garzón, A., Miklas, P. N., Lee, R., Clevenger, J., Myers, J., Korani, W., & McClean, P.E. (2023). Integrating de novo QTL-seq and linkage mapping to identify quantitative trait loci conditioning physiological resistance and avoidance to white mold disease in dry bean. *Plant Genome, 16*:e202380. https://doi.org/10.1002/tpg2.20380

Sarnoski, P.J., J.V. Johnson, K.A. Reed, J.M. Tanko, & S.F. O'Keefe. (2012). Separation and characterization of proanthocyanidins in Virginia type peanut skins by LC–MSN. *Food Chemistry, 131(3)*: 927–939. https://doi.org/10.1016/j.foodchem.2011.09.081

Scott, M. F., Ladejobi, O., Amer, S., Bentley, A. R., Biernaskie, J., Boden, S. A., Clark, M., Dell'Acqua, M., Dixon, L. E., Filippi, C. V., Fradgley, N., Gardner, K. A., Mackay, I. J., O'Sullivan, D., Percival-Alwyn, L., Roorkiwal, M., Singh, R. K., Thudi, M., Varshney, R.

K., Venturini, L., Whan, A., Cocram, J. & Mott, R. (2020). Multi-parent populations in crops: A toolbox integrating genomics and genetic mapping with breeding. *Heredity, 125(6)*, 396–416. https://doi.org/10.1038/s41437-020-0336-6

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.

Shirasawa, K., P. Koilkonda, K. Aoki, H. Hirakawa, S. Tabata, M. Watanabe, M. Hasegawa, H. Kiyoshima, S. Suzuki, C. Kuwata, Y. Naito, T. Kuboyama, A. Nakaya, S. Sasamoto, A. Watanabe, M. Kato, K. Kawashima, Y. Kishida, M. Kohara, A. Kuribayashi, C. Takahashi, H. Tsuruoka, T. Wada, & S. Isobe. (2012). In silico polymorphism analysis for the development of simple sequence repeat and transposon markers and construction of linkage map in cultivated peanut. *BMC Plant Biology*, *12(1)*: 80. https://doi.org/10.1186/1471-2229-12-80

Stange, M., Utz, H. F., Schrag, T. A., Melchinger, A. E., & Würschum, T. (2013). High-density genotyping: An overkill for QTL mapping? lessons learned from a case study in maize and simulations. *Theoretical and Applied Genetics, 126*(10), 2563–2574. https://doi.org/10.1007/s00122-013-2155-0

Sweeney, M. T., Thomson, M. J., Pfeil, B.E. & McCouch. S. (2006). Caught red-handed: RC encodes a basic helix-loop-helix protein conditioning red pericarp in Rice. *The Plant Cell*, *18*(2): 283–294. https://doi.org/10.1105/tpc.105.038430

Tallury, S.P., Isleib, T.G., Copeland, S.C., Rosas-Anderson, P., Balota, M., Singh, D., & Stalker, H.T. (2014). Registration of two multiple disease-resistant peanut germplasm lines derived from *Arachis cardenasii* Krapov. and WC Gregory, GKP 10017. *Journal of Plant Registration, 8,* 86-89. https://doi.org/10.3198/jpr2013.04.0017crg

Wickham, H. (2016). *Create elegant data visualizations using the grammar of graphics*. Create Elegant Data Visualizations Using the Grammar of Graphics. https://ggplot2.tidyverse.org/index.html

Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., Yuan, X., Zhu, M., Zhao, S., Li, X., & Liu, X. (2021). RMVP: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics, Proteomics & Bioinformatics, 19*(4), 619–628. https://doi.org/10.1016/j.gpb.2020.10.007

Zhou, Y., Yu, Z., Chebotarov, D., Chougule, K., Lu, Z., Rivera, L. F., Kathiresan, N., Al-Bader, N., Mohammed, N., Alsantely, A., Mussurova, S., Santos, J., Thimma, M., Troukhan, M., Fornasiero, A., Green, C. D., Copetti, D., Kudrna, D., Llaca, V., Lorieux, M., Zuccolo, A., Ware, D., McNally, K., Zhang, J., & Wing, R. A. (2023). Pan-genome inversion index reveals evolutionary insights into the subpopulation structure of Asian rice. *Nature Communications, 14*(1). https://doi.org/10.1038/s41467-023-37004-y.

Zhao, Z., Wu, M., Zhan, Y., Zhan, K., Chang, X., Yang, H. & Li, Z. (2017). Characterization and purification of anthocyanins from black peanut (*Arachis hypogaea* L.) skin by combined column chromatography. *Journal of Chromatography,* A 1519: 74–82. https://doi.org/10.1016/j.chroma.2017.08.078

Zhuang, W., H. Chen, M. Yang, J. Wang, M.K. Pandey, C. Zhang, W.-C. Chang, L. Zhang, X. Zhang, R. Tang, V. Garg, X. Wang, H. Tang, C.-N. Chow, J. Wang, Y. Deng, D. Wang, A.W. Khan, Q. Yang, T. Cai, P. Bajaj, K. Wu, B. Guo, X. Zhang, J. Li, F. Liang, J. Hu, B. Liao, S. Liu, A. Chitikineni, H. Yan, Y. Zheng, S. Shan, Q. Liu, D. Xie, Z. Wang, S.A. Khan, N. Ali, C. Zhao, X. Li, Z. Luo, S. Zhang, R. Zhuang, Z. Peng, S. Wang, G. Mamadou, Y. Zhuang, Z. Zhao, W. Yu, F. Xiong, W. Quan, M. Yuan, Y. Li, H. Zou, H. Xia, L. Zha, J. Fan, J. Yu, W. Xie, J. Yuan, K. Chen, S. Zhao, W. Chu, Y. Chen, P. Sun, F. Meng, T. Zhuo, Y. Zhao, C. Li, G. He, Y. Zhao, C. Wang, P.B. Kavikishor, R.-L. Pan, A.H. Paterson, X. Wang, R. Ming, & R.K. Varshney. (2019). The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nature Genetics, 51*(5): 865–876. https://doi.org/10.1038/s41588-019-0402-2

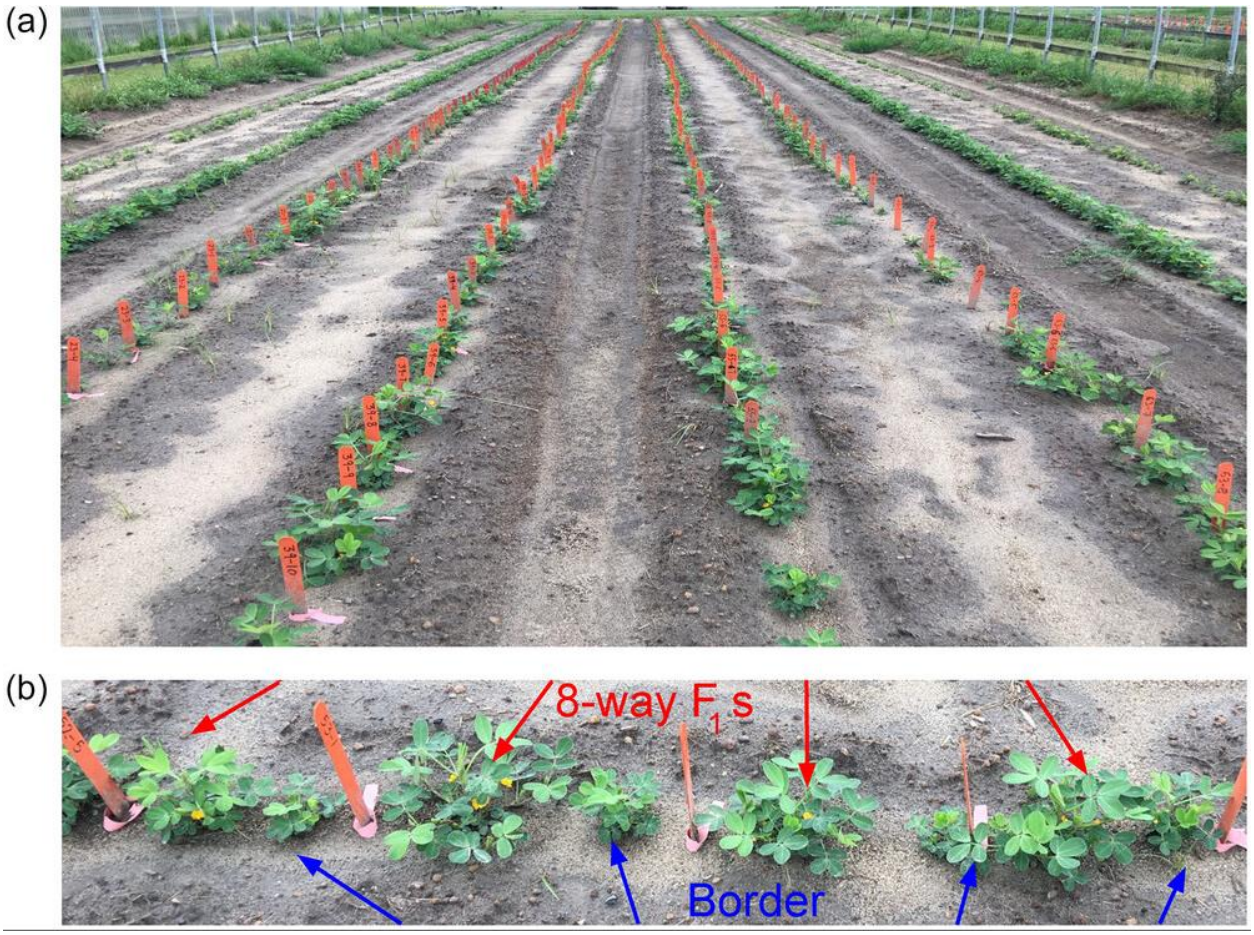


Figure 2.1 A multiparent advanced generation inter-cross (MAGIC) population of peanut (PeanutMAGIC) development and the 8-way F_1 plants. Close attention was given when advancing 8-way F_1 hybrid to produce F_2 seeds (A). Each F_1 seed was coded with two numbers, the pair number and the F_1 seed number produced by this pair. The line coded "39-9" indicates that this F_1 plant was the 9th hybrid F_1 seed produced by the pair No. 39. (B) The F_1 seeds individually planted with an alternating boarder of Georgia-06G (Branch, 2007), a popular commercial variety. The boarder plants were removed later to provide additional room for the F_1 plant to maximize the production of F_2 seeds. The hybrid vigor or heterosis was evidenced by the noticeably larger plants than the boarder plants.

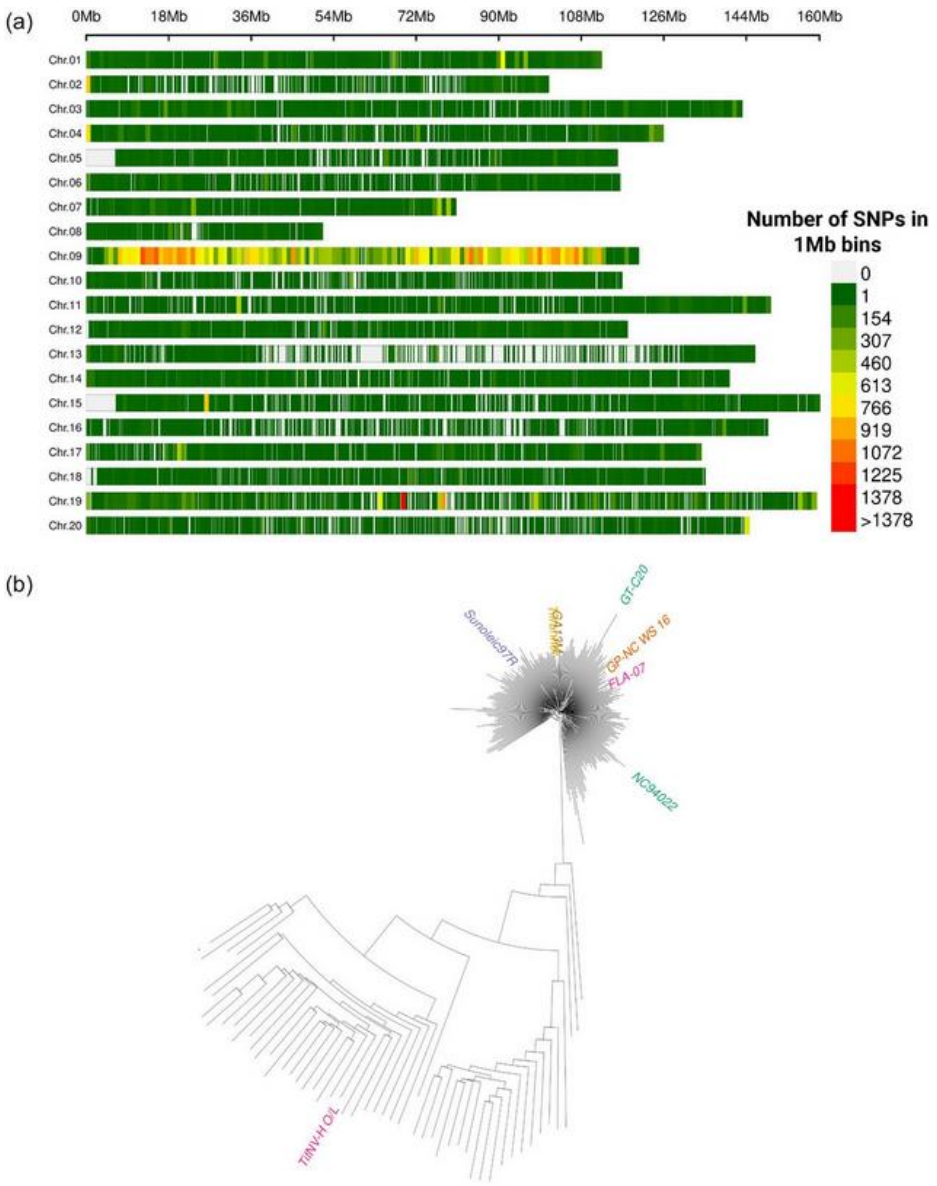


Figure 2.2 SNP marker distribution on the genome of PeanutMAGIC and Neighbor Joining (NJ) phylogenetic relationships among the MAGIC Core and eight founders. (A) Genome-wide distribution of 138,151 SNP markers on 20 chromosomes identified for the MAGIC Core. There was increased density in Chr. 09 and Chr. 19 harboring 65,687 and 12,045 SNPs, respectively, from the *A. cardenasii* introgression. (B) NJ tree of the MAGIC Core lines and founders using all SNP markers. There were 52 lines and the donor parent TifNV-H O/L as outliers, representing 16.8% of the population.

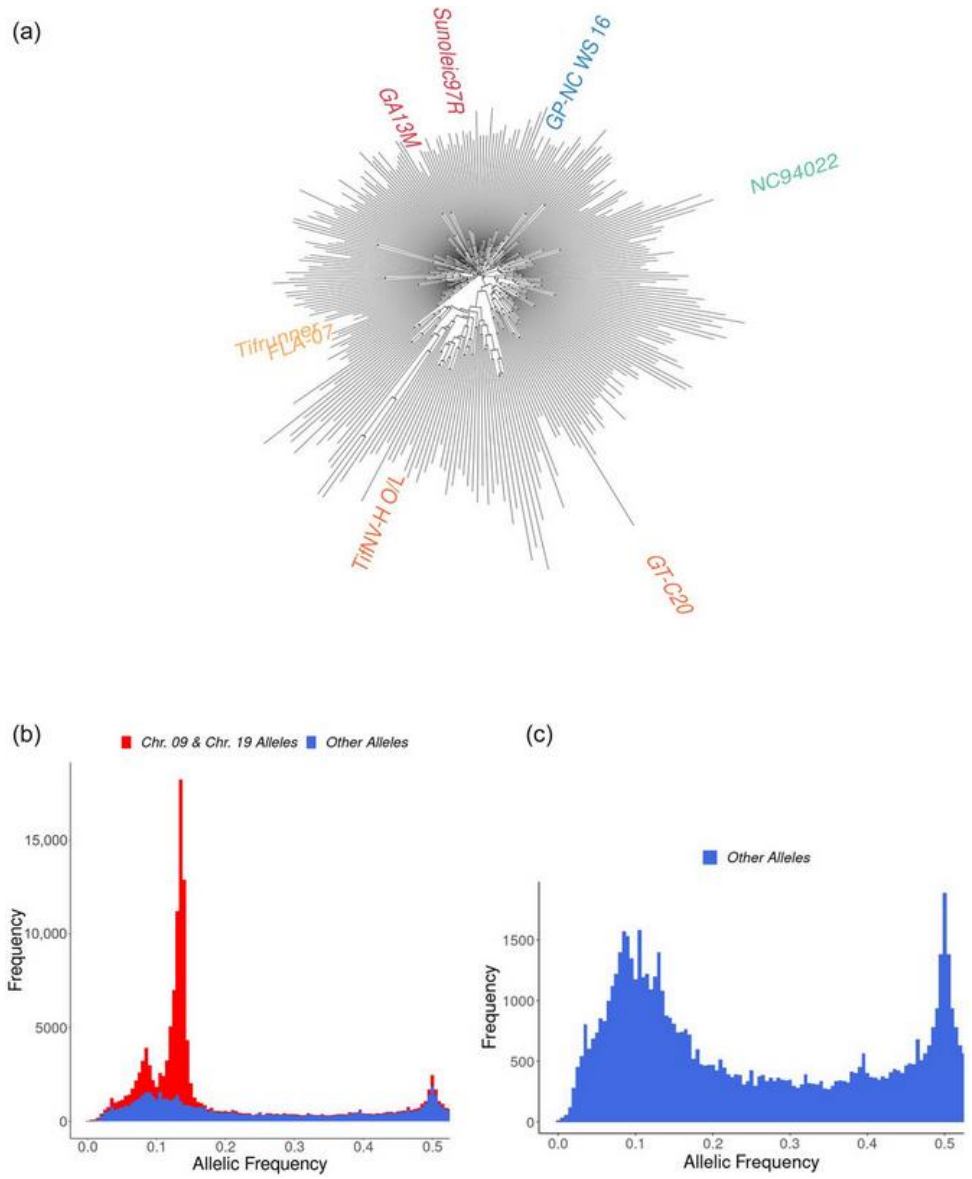


Figure 2.3 A phylogenetic tree and minor allelic frequency (MAF) of MAGIC Core. (A) The phylogenetic relationship of the RILs and founders using all SNPs excluding SNPs from Chr. 09 and Chr. 19. (B) MAF distribution using all SNPs, peaking at about 0.12 or 12% throughout the population. Alleles in red from Chr. 09 and Chr. 19, harboring excessive diversity from the introgression of diploid peanut. (C) MAF distribution using all alleles excluding Chr. 09 and Chr. 19, peaking at 0.1 or 10% suggesting PeanutMAGIC was a diverse synthetic population with near theoretical proportions of founder haplotypes.

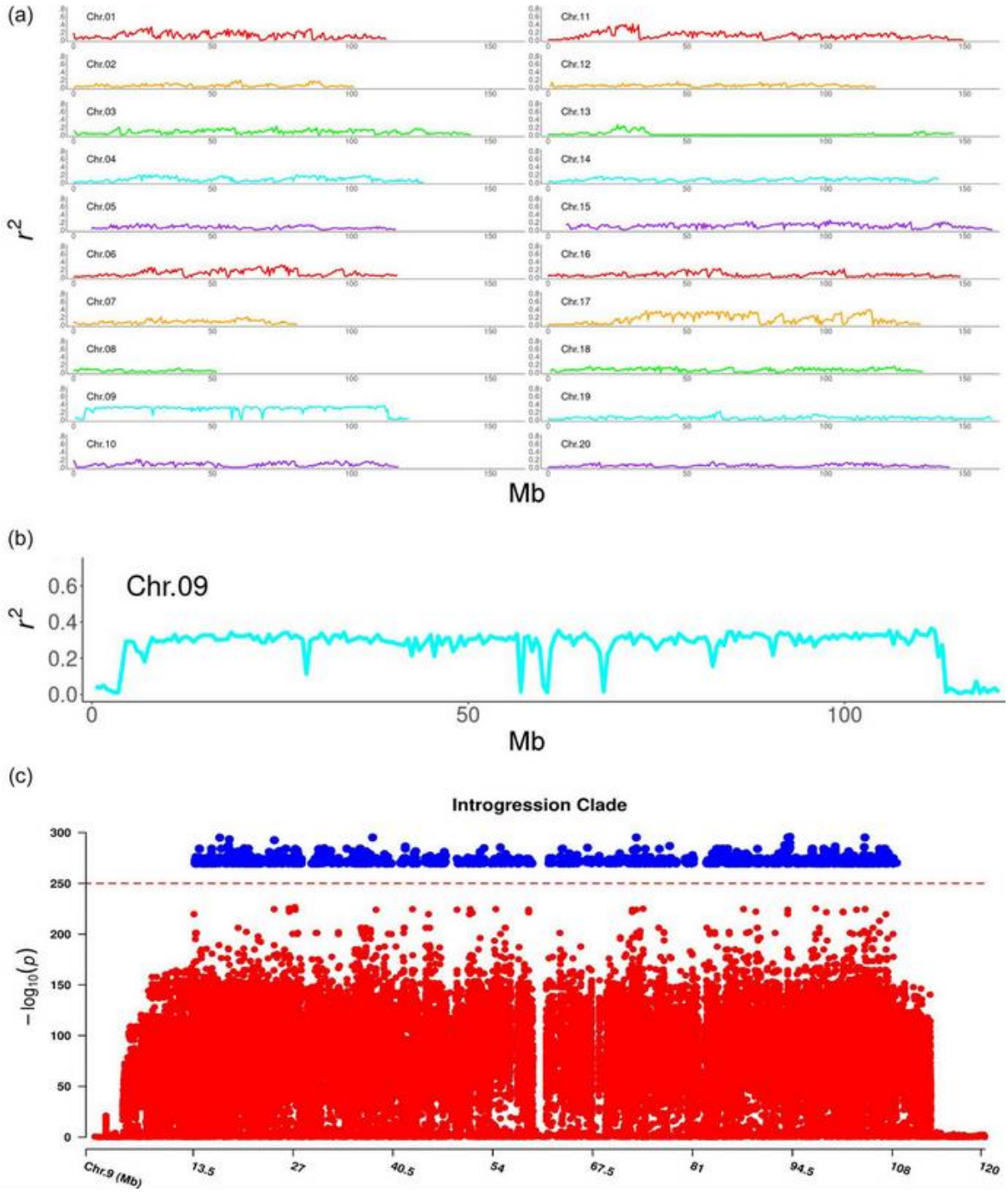


Figure 2.4 Genomic linkage disequilibrium (LD) along 20 peanut chromosomes in PeanutMAGIC population and a Manhattan plot of genotypes with introgression recombinants in Chr. 09. (A) The r^2 values of markers calculated in a sliding 100 marker window, averaged in 500 kb bins, and plotted against physical location on the corresponding chromosome. Low r^2

values indicate low LD or high recombination rate in this region and *vice versa*. (B) The r^2 values along Chr. 09, showing higher r^2 values (plateaus) with lower recombination in the regions of *A. cardenasii* introgression (Chu et al. 2016; Clevenger et al., 2017), however, there were regions where high recombination rate did occur as denoted by the "valleys" across the r^2 plot. (C) These locations align with the clades in the phylogenetic trees (Figure 2B) and the outgroup. The increase in -log (*p* value) (in blue) shows locations the MAGIC Core introgression lines share, where -log (*p* value) was lower indicating the clade with specific introgression recombination points (in red).

Figure 2.5 Morphological variation in pod and seed characteristics of two unique 8-way crosses and their derived $F_{2:3}$ families. (A) One F_1 seed from pair No. 50, coded as 50-10, produced 12 $F_{2:3}$ families (individual groups) highlighting the morphologic differences in pod shapes including pod constriction, reticulation, and the number of seeds per pod within one cross. (B) Pair No. 43 produced 20 F_1 hybrid seeds and four $F_{2:3}$ families were selected from four different F_1 seeds (coded as 43-4, 43-5, 43-6, and 43-20) with different pod and seed characteristics. New traits in pod and seed, such as dark seed coat color and 3-seeded pod, have been produced. None of these traits were present in the founder lines.

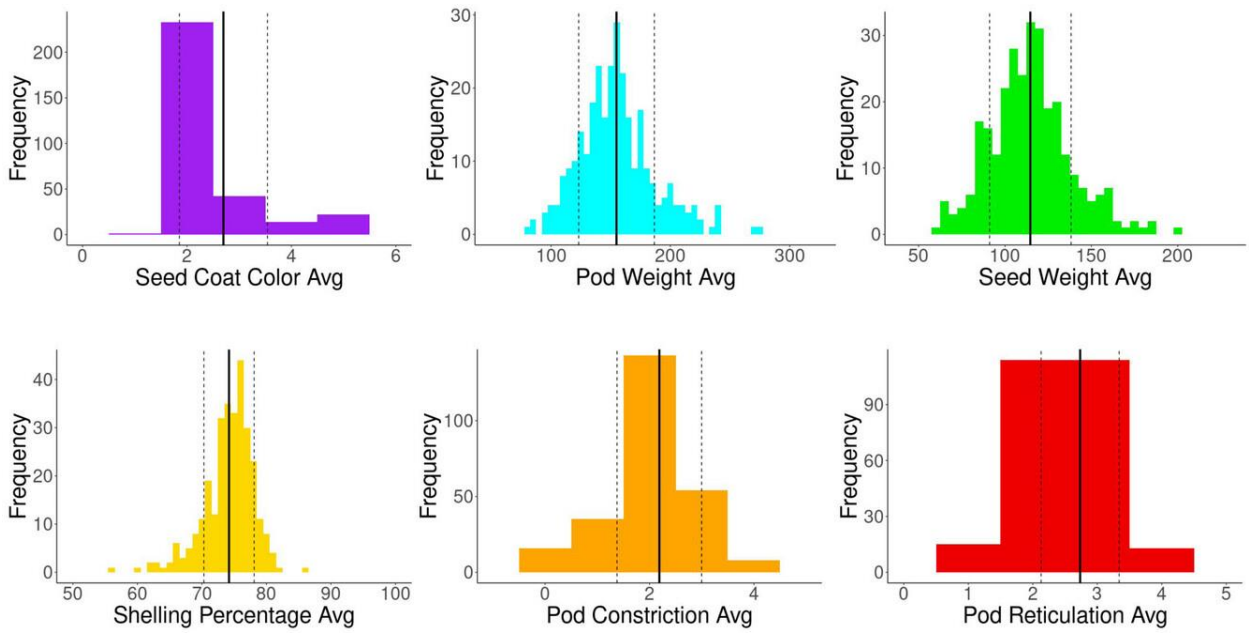


Figure 2.6 Histogram plots represent the average scores of two-year phenotypes for seed coat color (SCC in purple), pod weight (PW in blue), seed weight (SW in green), shelling percentage (SP in yellow), pod constriction (PC in orange), and pod reticulation (PR in red) in the MAGIC Core. Solid vertical lines represent the average value of the phenotype. Dashed vertical lines represent the average plus or minus one standard deviation.

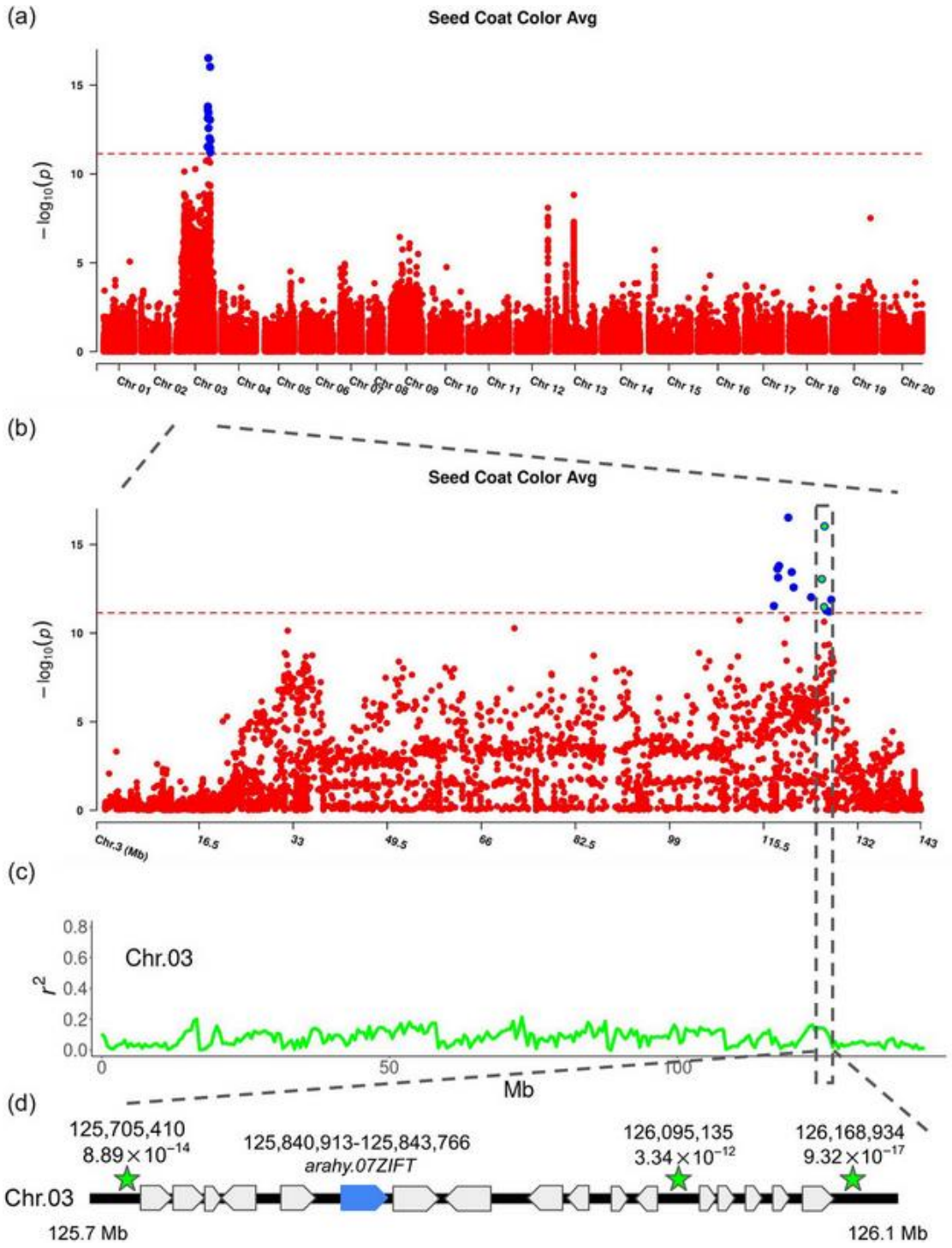


Figure 2.7 Manhattan plot of a genome-wide association study (GWAS) on peanut seed coat color (SCC) and identification of the candidate genes. (A) Manhattan plot of GWAS on SCC and chromosome-wide significance level set at threshold –log10 (*p*) of 12. The dashed red colored

line was the significant threshold for SNPs. (B) A zoomed view of Chr. 03 with the significant associated SNPs located in the region from 117.3 - 127.3 Mb. (C) The dashed box shows a narrowed genome region from 125.7 - 126.1 Mb including three consistent SNPs (in green) in panel B, which had elevated LD with higher r^2 values. (D) Zoomed in view of Chr. 03 with the two dashed lines shows the locations of three SNP markers (in green stars) in the flanking regions to a putative candidate gene *arahy.07ZIFT* (in blue).

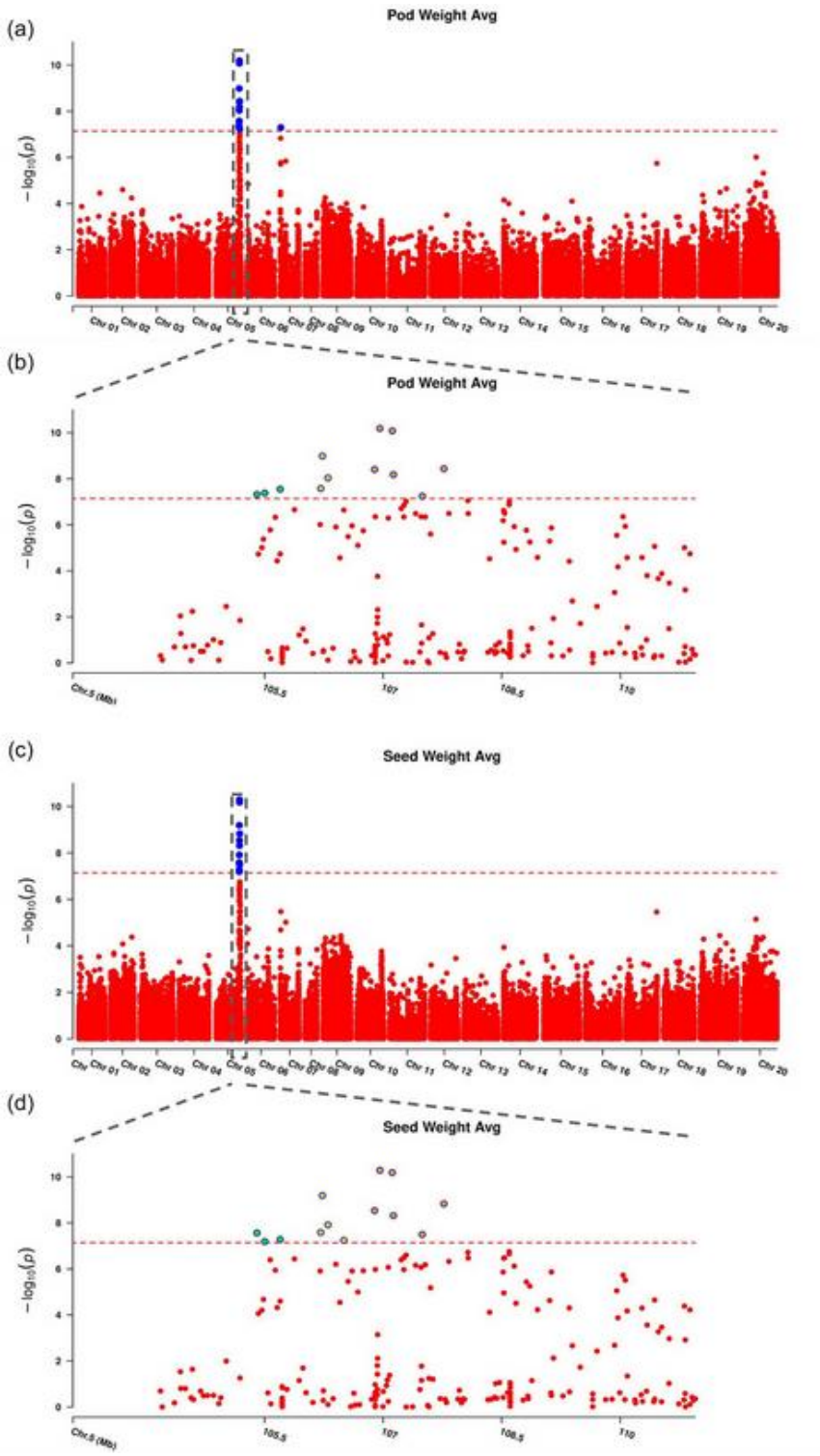


Figure 2.8: Manhattan plot of a genome-wide association study (GWAS) on peanut pod weight (PW) and seed weight (SW). (A) Manhattan plot of PW shows two significant locations above

the Bonferroni threshold of 7 on Chr. 05 and Chr. 07, respectively. (B) Zoomed in view of Chr. 05 from 104 - 111 Mb identified three distinct significant regions above the Bonferroni threshold, as highlighted in green, yellow, and orange, respectively. (C) Manhattan plot of SW shows a significant region on Chr. 05, colocalized with PW. (D) Zoomed in view of Chr. 05 in the same region as PW identified three distinct regions above the threshold as PW, as highlighted in green, yellow, and orange, respectively.

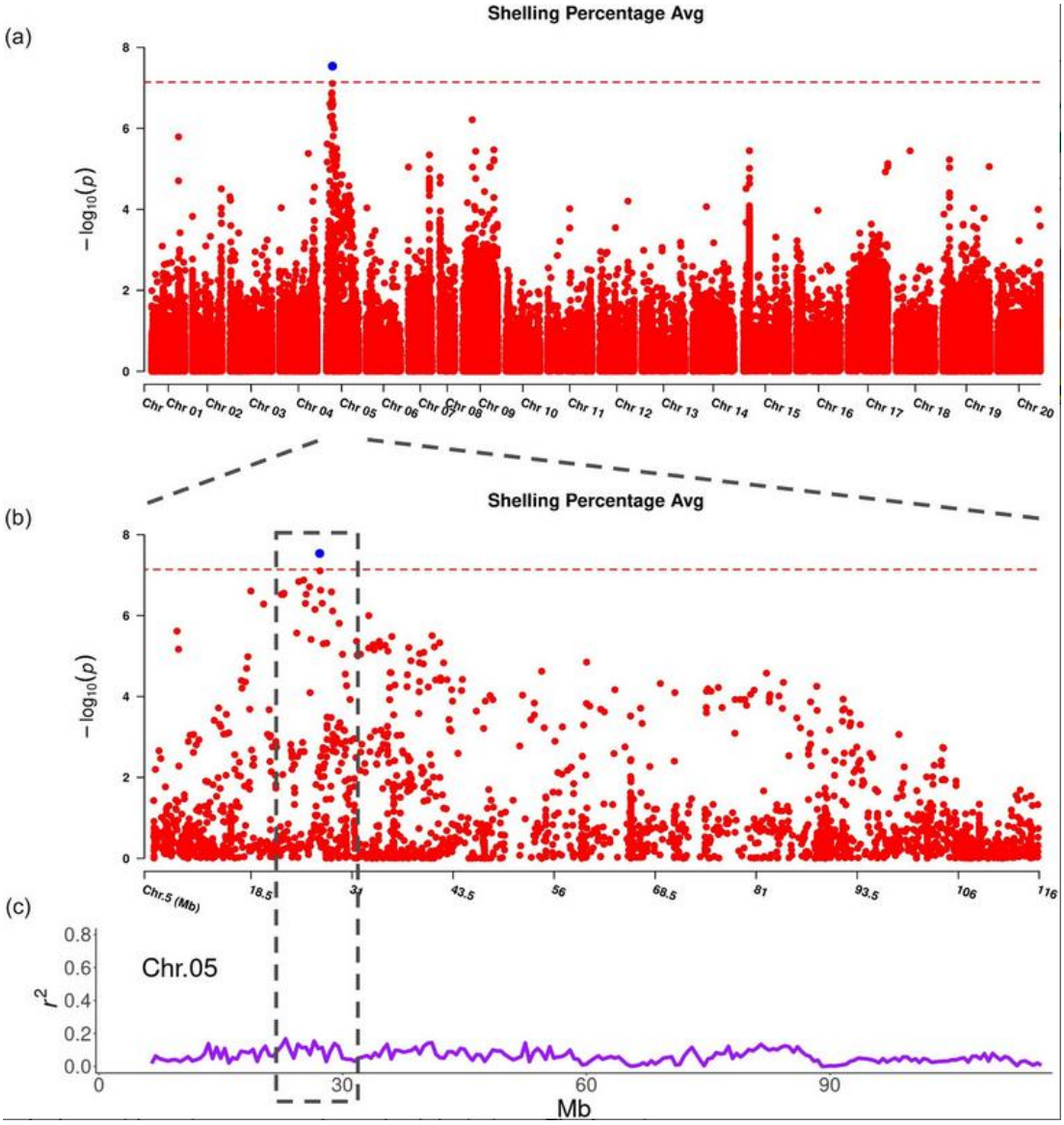
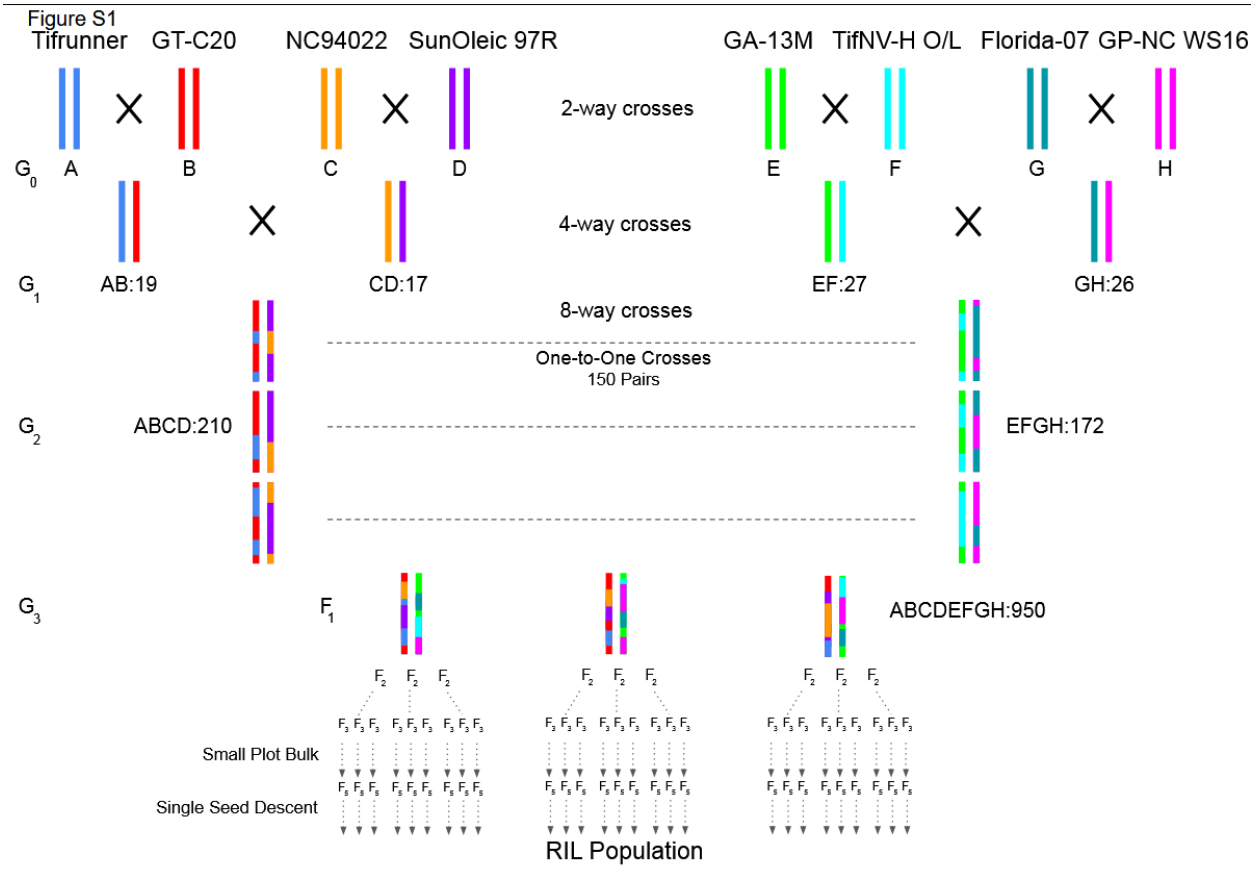


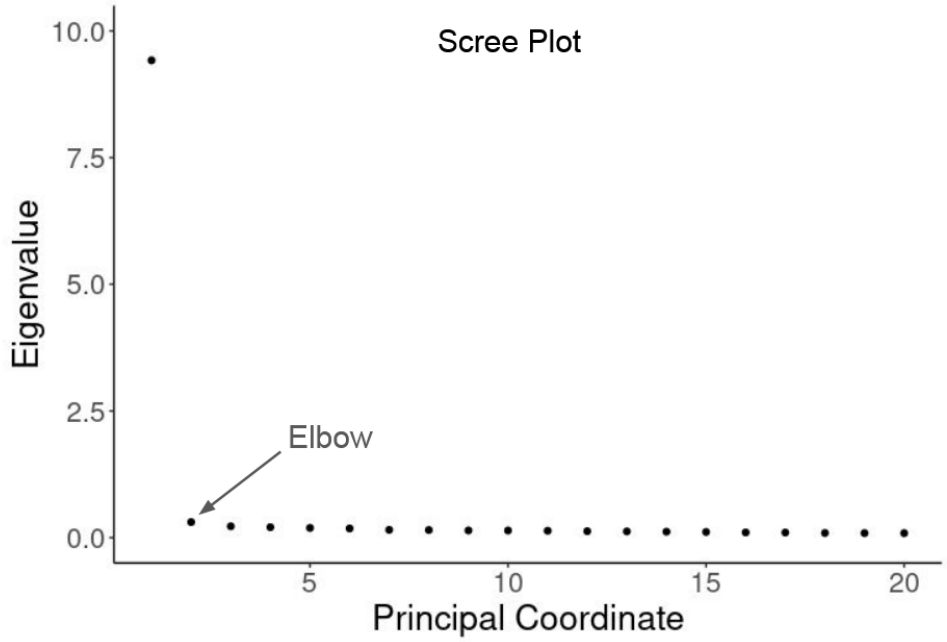
Figure 2.9 Manhattan plot of a genome-wide association study (GWAS) on peanut shelling percentage (SP) and the linkage disequilibrium (LD) block on Chr. 05. (A) Manhattan plot of SP shows significant SNPs above the Bonferroni threshold on Chr. 05. (B) Zoomed in view of Chr. 05 shows the elevated –log10 (*p*) across the chromosome. (C) The r^2 plot of Chr. 05 shows the low LD throughout the population in the associated region (dashed box), suggesting recombination is only occurring within high SP haplotypes in comparison with panel B with higher LD.



Supplementary Figure 2.1 PeanutMAGIC population breeding scheme. Eight peanut cultivars and breeding lines were crossed in a simple funnel inter-cross breeding design as [A×B/C×D]/[E×F/G×H] through three generations of advanced inter-crossing and selfing to produce 3,187 PeanutMAGIC RILs at the F_7 generation. Letters and colors refer to the founders as indicated at the top. Colored bars in the middle depict the recombination of a single chromosome through each generation, where each chromosome represents an individual hybrid. At G_0 the eight inbred founders were crossed with one other founder in a simple funnel design. The 2-way hybrids (G_1) (19 AB, 17 CD, 27 EF, and 26 GH, respectively) were crossed to generate 4-way hybrids (G_2) (210 ABCD and 172 EFGH, respectively). 4-way hybrids were crossed from one plant to another plant only, generating 950 8-way hybrids from 150 successful

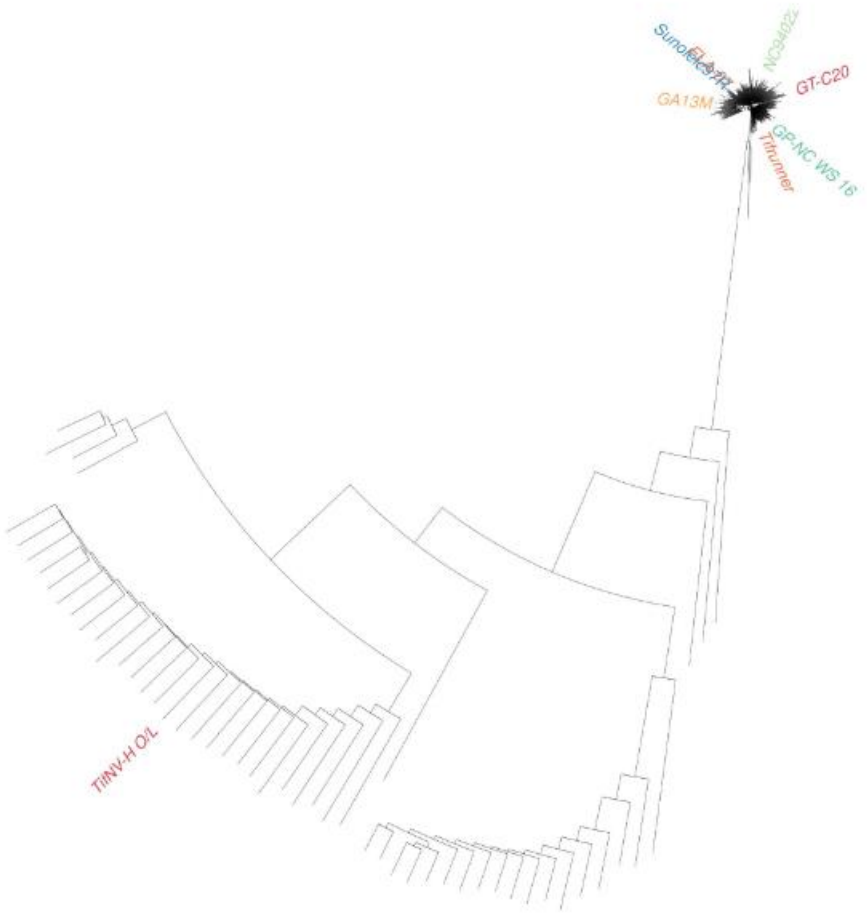
pairs of parents. The 8-way hybrids were self-pollinated to generate $F_{2:3}$ families. The $F_{2:3}$ families were advanced to $F_{2:7,}$ resulting in 3,187 $F_{2:7}$ recombinant inbred lines (RILs).

Figure S2

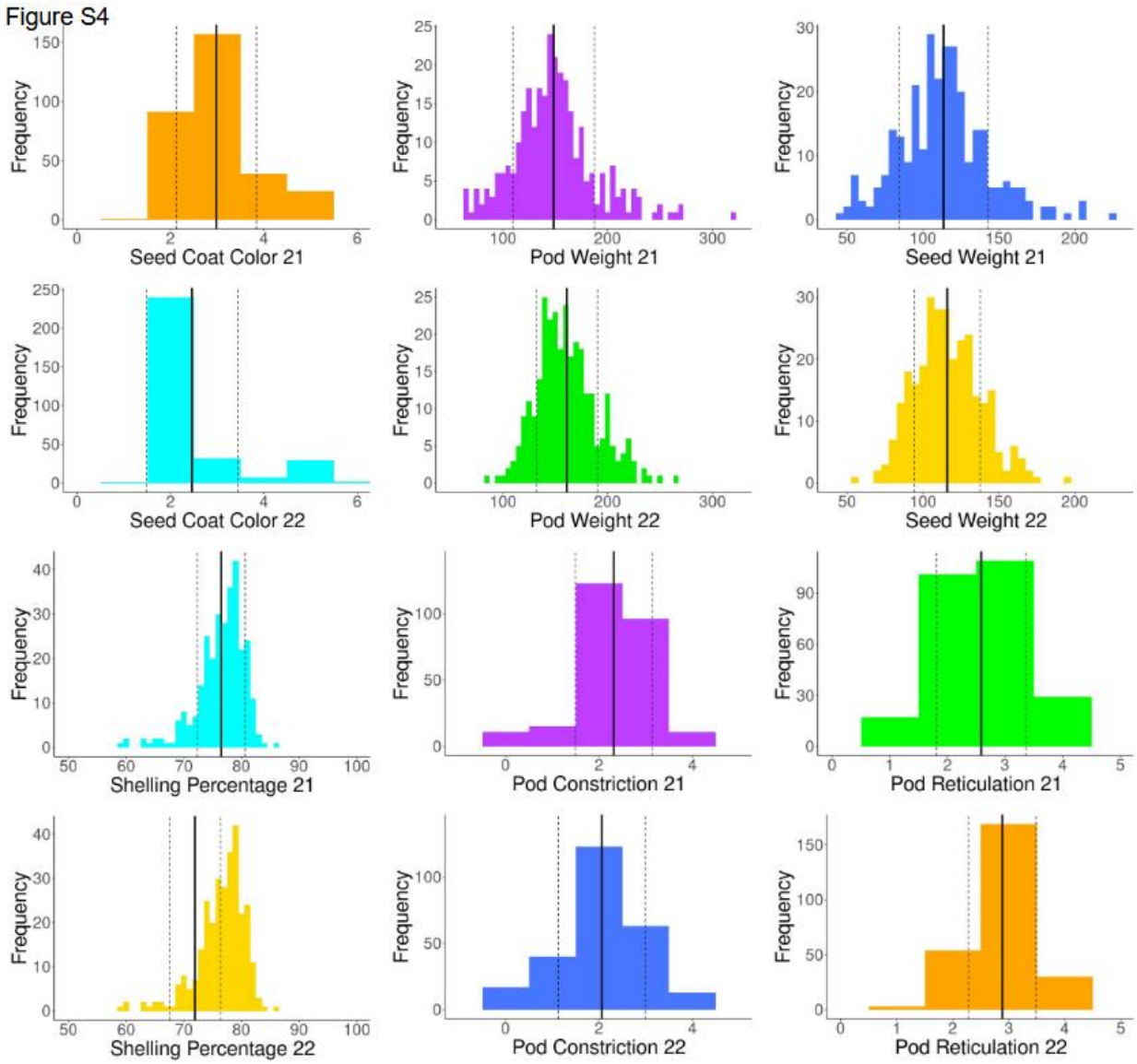


Supplementary Figure 2.2 Scree plot. Axes of coordinates calculated through multidimensional scaling (MDS) plotted against respective eigenvalues. The elbow of the scree plot is highlighted to determine the number of axes to control population structure.

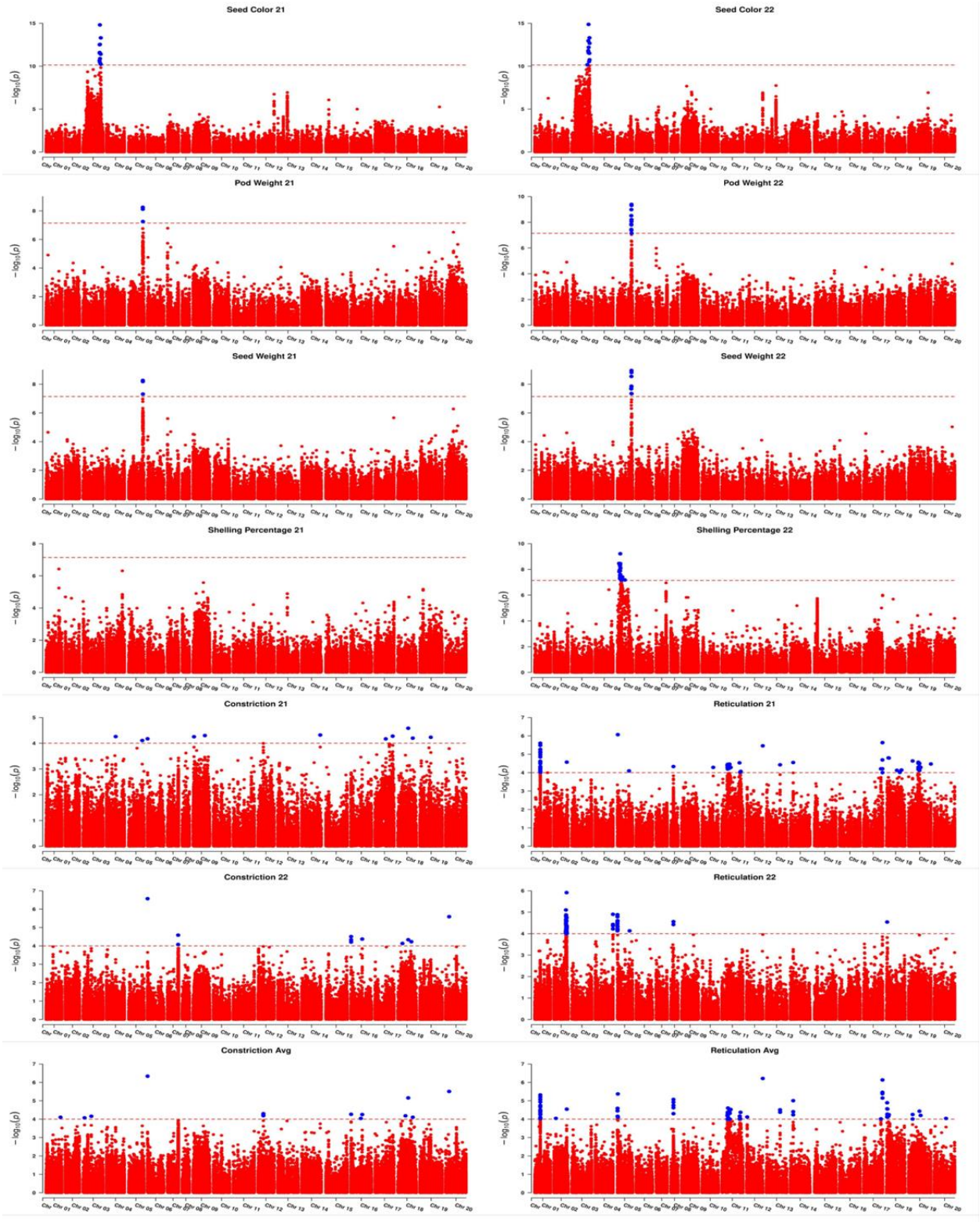
Figure S3



Supplementary Figure 2.3 Neighbor joining phylogenetic tree using SNP markers only on Chr. 09 and Chr. 19.

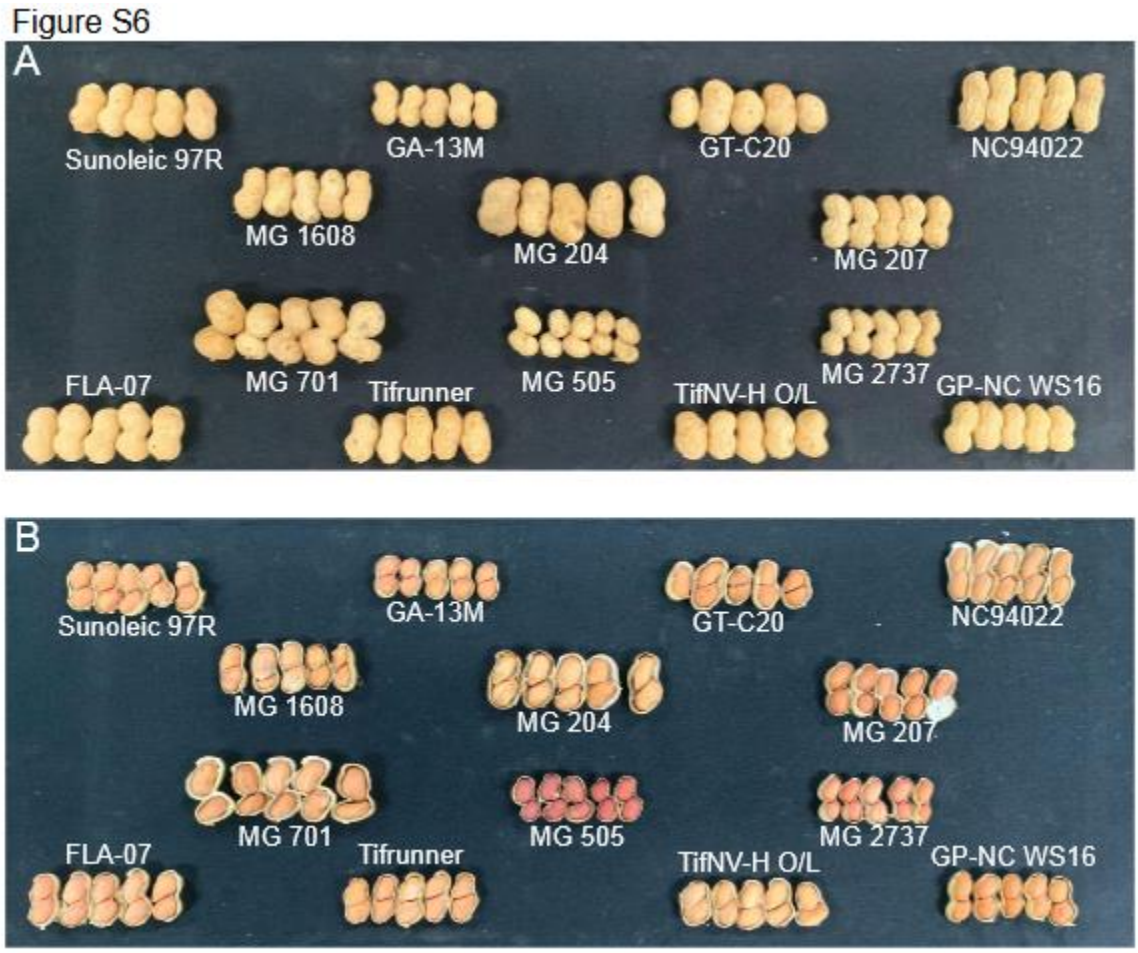


Supplementary Figure 2.4 Histograms of peanut seed coat color (SCC), pod weight (PW), seed weight (SW), shelling percentage (SP), pod constriction (PC), and pod reticulation (PR) phenotypes for two years, 2021 and 2022.



Supplementary Figure 2.5 Manhattan plots of genome-wide association study (GWAS) on peanut seed coat color (SCC), pod weight (PW), seed weight (SW), shelling percentage (SP),

pod constriction (PC), and pod reticulation (PR) phenotypes of 2021 and 2022, respectively. The averages of PR and PC plots are presented at the bottom.



Supplementary Figure 2.6 Morphological diversity of pod and seed traits in MAGIC Core. (A) Pods of all eight founders and select PeanutMAGIC lines illustrated the diversity in pod reticulation (PR), pod constriction (PC), and pod size. (B) Same as Panel (A) demonstrated the seed coat color (SCC), seed size, and shell thickness phenotypes.

Phenotype	Min	Max	Average	Standard Deviation	Seed Coat Color scale		
Seed Coat Color 2021	1	5	2.98	0.86	0 = no color	3 = pink	6 = dark purple
Seed Coat Color 2022	1	6	2.46	0.98	1 = white	4 = red	
Seed Coat Color Average	1	5.5	2.70	0.84	2 = tan	5 = purple	
Pod Weight 2021 (g)	62.77	320.50	148.57	39.05			
Pod Weight 2022 (g)	86.67	265.43	161.07	29.04			
Pod Weight Average (g)	78.69	273.48	154.82	31.69			
Seed Weight 2021 (g)	46.39	225.70	113.46	29.43			
Seed Weight 2022 (g)	54.29	197.29	115.84	21.91			
Seed Weight Average (g)	58.23	201.33	114.65	23.58			
Shelling Percentage 2021 (%)	59.04	86.49	76.48	4.16			
Shelling Percentage 2022 (%)	49.41	91.17	71.92	4.43			
Shelling Percentage Average ('	55.96	85.83	74.20	3.89			
Pod Constriction 2021	0	4	2.32	0.82	Pod Constriction scale		
Pod Constriction 2022	0	4	2.06	0.93	0 = none	2 = moderate	4 = very deep
Pod Constriction Average	0	4	2.19	0.81	1 = slight	3 = deep	
Pod Reticulation 2021	1	4	2.59	0.78	Pod Reticulation scale		
Pod Reticulation 2022	1	4	2.88	0.60	1 = smooth	3 = moderate	
Pod Reticulation Average	1	4	2.73	0.61	2 = slight	4 = rough	

Supplementary Table 2.1 MAGIC Core phenotypic statistic data for seed coat color (SCC), pod weight (PW), seed weight (SW), shelling percentage (SP), pod constriction (PC), and pod reticulation (PR) for year-1 (2021), year-2 (2022), and the average. Scales for SCC, PC, and PR are included. PW, SW, and SP are in grams.

Line	Seed Coat Color	Pod Weight Avg (g)	Seed Weight Avg (g)	Shelling Percentage Avg (%)	Constriction	Reticulation	Seed coat color scal	Pod Constriction sca	Pod Reticulation scale
MG105	2	116.39	77.39	66.96	1	3	0 = no color	0 = none	1 = smooth
MG1102	2	111.31	89.83	80.83	1	3	1 = white	1 = slight	2 = slight
MG1203	2	97.10	71.25	73.64	0	2	2 = tan	2 = moderate	3 = moderate
MG1206	4	116.71	89.54	77.05	0	1	3 = pink	3 = deep	4 = rough
MG1210	2	110.25	86.74	80.28	2	1	4 = red	4 = very deep	
MG1405	2	189.67	146.36	77.26	2	3	5 = purple		
MG16	2	111.81	87.95	78.74	2	1	6 = dark purple		
MG1608	2	143.76	122.44	85.83	2	3			
MG1609	5	171.04	129.43	75.67	4	2			
MG1807	6	153.67	108.19	69.66	4	2			
MG204	2	233.28	183.93	78.61	1	2			
MG207	2	140.32	107.34	76.36	3	4			
MG2208	2	105.26	80.99	77.12	2	3			
MG2210	2	98.08	68.09	69.60	1	1			
MG2604	3	122.27	82.21	66.20	0	4			
MG2606	2	192.59	148.74	77.31	2	2			
MG2737	2	106.70	77.22	74.34	3	2			
MG403	2	156.95	112.51	71.67	4	4			
MG404	2	165.06	127.11	77.47	0	4			
MG409	5	96.54	65.78	67.99	3	3			
MG505	5	154.00	119.14	77.37	4	3			
MG507	1	136.80	84.98	62.18	2	3			
MG601	3	112.95	82.50	73.13	2	2			
MG609	2	139.66	112.70	80.88	2	2			
MG701	2	268.20	201.33	75.06	4	3			
MG801	2	83.51	66.92	63.61	1	2			

Supplementary Table 2.2 MAGIC Core lines exhibiting various combinations of traits. All phenotypes represent the average of two years data. A selection of 26 PeanutMAGIC lines are described, which express various phenotypes including seed coat color, small seeded, large seeded, thick-shelled, thin-shelled, low constriction and low reticulation, available for breeding programs..

CHAPTER 3

PEANUTMAGIC MULTIPARENT POPULATION-SPECIFIC PANGENOME UNVEILS A NOVEL THIRD FAD2 GENE AND SOLVES THE MYSTERY OF MID-OLEIC FATY ACID IN PEANUT

[1] Thompson, E., Korani, W., Wu, D., Garg, V., Tonnis, B., Wang, M., Holbrook, C. C., Ozias-Akins, P., Culbreath, A. K., Varshney, R. K., Guo, B., & Clevenger, J. (2025). PeanutMAGIC multiparent population-specific pangenome unveils a novel third FAD2 gene and solves the mystery of mid-oleic fatty acid in peanut. Submitted to *Nature Communications*, 2/10/2025.

Abstract

Cultivated peanut (*Arachis hypogaea* L.) is an important oil and protein crop. However, mid-oleic fatty acid peanuts have puzzled the peanut research community, using the traditional two-gene model. Here, we present a population-specific pangenome using the eight founder genomes of the PeanutMAGIC population. This graph-based pangenome serves as a comprehensive reference, capturing all segregating founder variations within the PeanutMAGIC population of 3,187 recombinant inbred lines (RILs). The whole genome sequencing was conducted for the MAGIC Core, a subset of 310 RILs, to generate genotypes. Using the genotyping data derived from the population-specific pangenome, we visualized genomic recombination within the MAGIC Core for detailed marker-trait association analysis for oleic acid content. This investigation identified a novel third fatty-acid desaturase 2 (*AhFAD2*) gene, named *AhFAD2C*, located on the same chromosome as *AhFAD2B*. When recombination occurs, *AhFAD2C* can segregate from *AhFAD2B* and mid-oleic genotype is revealed, resolving the long-standing mystery of a three-gene model theory in high-oleic peanuts. Our findings underscore the limitations of using a single-reference genome, which can lead to false association in marker discovery and downstream applications. In contrast, a population-specific pangenome provides a more accurate and reliable framework for genomic studies. This approach offers new insights into peanut oil quality and demonstrates the advantages of using a multiparental population-specific pangenome for complex trait analysis.

Introduction

Arachis hypogaea L. is cultivated globally as a sustainable and affordable oil and protein source, yielding 54 million tons annually (http://www.fao.org/faostat, 2020). Peanut seeds

contain considerable amounts of oil (~50%, dry weight) and protein (~25%, dry weight) that are bioavailable and beneficial for human health (Jones et al., 2014; Kris-Etherton et al., 2008; Alper & Mattes 2003). Oleic and linoleic fatty acids are two heart-healthy, major components of the total oil content in peanut seeds. Therefore, peanut seed oil composition is an important quality trait for confectionery and wholesale markets, leading to a focus on oil composition in breeding programs and functional studies (Davis et al., 2016; Chu et al., 2011). High-oleic (>75%) peanut cultivars are desirable for their increased shelf life and improved palatability in whole nuts or as ingredients in peanut products, whereas low (<55%) to mid-oleic (55-75%) cultivars, with high-palmitic acid content, are valued for peanut butter-based products. In cultivated peanut, fatty acid desaturase 2 (AhFAD2) is the enzyme responsible for the conversion of oleic acid into linoleic acid, with the functional genes *AhFAD2A* and *AhFAD2B* (Jung et al., 2000; López et al., 2000). Loss-of-function mutations in *AhFAD2A* and *AhFAD2B* have been identified in breeding populations, with selectable markers available for marker-assisted selection of high-oleic phenotypes (Chu et al., 2007; Chu et al., 2009; Chu et al., 2011). The *AhFAD2A* and *AhFAD2B* markers have been successfully utilized in the development of a high-oleic and root-knot nematode-resistant variety, 'TifNV High O/L', suggesting that information of two genes is sufficient to predict and select for high-oleic varieties (Chu et al., 2011). However, mid-oleic peanut lines have been observed, but their inheritance could not be explained solely by the previously established two-gene model (Branch et al., 2022), leading to the hypothesis that a third recessive gene, along with *AhFAD2A* and *AhFAD2B*, is needed to maintain the high-oleic fatty acid trait in peanut (Branch et al., 2022). Interestingly, Pandey et al. (2014) reported that *AhFAD2B* had a higher phenotypic effect on oleic acid content than *AhFAD2A* using two biparental populations, and the normal distribution of oleic fatty acid in the mapping populations

did not fit the two-gene model (Pandey et al., 2014). The mystery of mid-oleic has perplexed the peanut community, particularly concerning the confectionery industry's need for consistent quality in products.

Biparental populations have been the primary focus of plant genetics. However, they are limited by their construction from a single cross involving only two genetic donors (Scott et al., 2020). Multiparental populations, such as multiparent advanced generation intercross (MAGIC) populations, address these limitations by incorporating multiple hybrid crosses and genetic donors (Thompson et al., 2024). This approach facilitates a segregating population with equal genetic contribution from all founders and known pedigrees, thereby enhancing diversity and fine recombination. The increased genetic diversity and recombination in MAGIC populations enable finer genetic mapping of traits such as oleic acid content to resolve the mystery that has eluded the peanut community. The PeanutMAGIC population exemplifies the advantages of a multiparental breeding approach in peanut, providing increased genetic variation and recombination for genomic studies (Thompson et al. 2024).

The community reference genome of *A. hypogaea* cv. 'Tifrunner' (TR) was a joint effort of the International Peanut Genome Sequencing Consortium (Bertioli et al. 2019). The publication of a reference genome signified a major milestone in the peanut community, due to the allotetraploid (AABB-type genome; 2n = 4x = 40 chromosomes) nature of this species and limited subgenome divergence that stemmed from the recent hybridization event from its related progenitors (*A. duranensis*, AA; *A. ipaensis*, BB). These factors have made genome sequencing and assembly challenging for tetraploid peanut (Bertioli et al., 2016; Bertioli et al., 2019). The reference genome has been pivotal for divulging the intricacies and origin of cultivated peanut for crop improvement; however, it does not reveal all the genetic variance within cultivated

peanut or a particular mapping population. Recent advancement in sequencing and bioinformatics has been driving a paradigm shift from a single reference-based genome to graph-based pangenome reference discovery in structural variants and trait mapping (Danilevicz et al., 2020; Guo et al., 2025). Conversely, a species level pangenome offers a library of all variations present throughout the species, more variants than within a specific population. The founders of the PeanutMAGIC population do not fully represent the global diversity of peanut. Therefore, the use of a species level pangenome might include variations that are not possible in the population, potentially introducing error and leading to improper genotyping for PeanutMAGIC recombinant inbred lines (RILs).

Here, we present chromosome-level assemblies for all the founders of PeanutMAGIC (Thompson et al. 2024) and constructed a population-specific pangenome for the identification of all inherited haplotypes for genotyping and uncovering putative causal variation. With considerable detail, we can detect recombination points that have been accumulated throughout the population synthesis using pangenome-based markers to identify the cause of the mid-oleic phenotype that puzzled the peanut research community (Supplementary Fig. 3.1).

Results

Genome assemblies of PeanutMAGIC founders and pangenome construction

The PeanutMAGIC founders, ‘Georgia-13M’ (13M) (Branch, 2014), ‘SunOleic 97R’ (97R) (Gorbet and Knauft, 2000), ‘GT-C20’ (C20) (Liang et al., 2005), ‘Florida-07’ (F07) (Gorbet and Tillman, 2009), ‘NC94022’ (NC) (Culbreath et al., 2005), and ‘TifNV-High O/L’ (TNV) (Holbrook et al., 2017), were sequenced using two cells of PacBio Sequel II, and ‘GP-NC WS16’ (WS16) (Tallury et al., 2014) was sequenced using one cell of PacBio Revio. Four of the

founders (13M, 97R, C20, and F07) were scaffolded using high-throughput chromosome conformation capture (Hi-C) data to generate chromosome-length assemblies (Supplementary Table 3.1). Three founders (NC, TNV, and WS16) were scaffolded using RagTag (Alonge et al., 2022), with the eighth founder (TR) as the reference (Supplementary Table 3.2). The genome size of the founders ranged from 2,490 Mb (F07) to 2,629 Mb (TNV) with contig N50 ranging from 21 Mb (F07) to 60 Mb (TNV) (Supplementary Table 3.3). The contig N50 for TR (Bertioli et al. 2019) was 1 Mb, highlighting the improved assembly quality achieved for the founder genomes. The variance of contig N50 for sequenced founders may stem from sequence quality differences and scaffolding efficiency between genotypes. The completeness of the assemblies was estimated through Benchmarking Universal Single-Copy Orthologs (BUSCO) (Manni et al., 2021), ranging from 95.2 (F07) to 97.6 (TR and WS16) (Supplementary Table 3.4). Notably, the duplicated BUSCO values among the founders ranged from 89.3 (13M and F07) to 93.3 (TR), indicating most of the allopolyploid gene duplications were maintained from the polyploidization event that led to present day peanut (Fig. 3.1a).

A graph-based pangenome was constructed using the Minigraph-Cactus pipeline with TR as the reference genome. This population-specific pangenome of PeanutMAGIC can be directly utilized with TR-based annotations (Hickey et al., 2023; Armstrong et al., 2020). The resulting acyclic, directed graph enables the resolution of founder haplotypes as traversed paths, with bubbles representing sites of variation, thereby preserving the founder origin of genetic variation. Within the PeanutMAGIC pangenome, 2,762,166 variant sites were identified, consisting of 1,606,159 SNPs and 1,156,007 non-SNP variants, including indels, inversions, and complex variants. Previously, only 138,151 markers were identified for PeanutMAGIC RILs, indicating that there are nearly 20 times more variants present in the population compared to the observed

markers used to evaluate the population (Thompson et al., 2024). To examine similarities within the allotetraploid genome, syntenic blocks were identified and show colinear sequences for homeologous and non-homeologous regions (Fig. 3.1b). The high collinearity between and within the A and B subgenomes supports previous studies and exemplifies the difficulties of generating markers in the past (Clevenger and Ozias-Akins, 2015; Pandey et al., 2017; Clevenger et al., 2018; Korani et al., 2021). Furthermore, syntenic block connections among paralogs underscore the duplication of genes within subgenomes, further complicating marker identification and genome assembly. The total number of variants per chromosome ranged from 34,463 (Chr.08) to 669,133 (Chr.09) (Fig. 3.1c). The increased variation on Chr.09 is due to the introgression from the wild diploid peanut *A. cardenasii* by the founder TNV (Thompson et al., 2024). The difference between the number of variants on homeologous chromosomes ranged from 989 (628 SNPs; 361 non-SNP; Chr.04 and Chr.14) to 153,392 (30,288 SNP; 123,104 non-SNP; Chr.01 and Chr.11), excluding Chr.09 and Chr.19, suggesting different levels of conservation between homeologs (Fig. 3.1c, Supplementary Table 3.5).

In the case of a population-specific pangenome, other published *A. hypogaea* genomes are available that can be leveraged to study genomic variations at a species level, as in other crops like *Vitis vinifera* or *Brassica oleracea* (Liu et al. 2024; Li et al. 2024), or at the genus level, such as *Citrullus* or *Cicer* (Zhang et al. 2024; Khan et al. 2024). As a reference library, a population specific pangenome should detail heritable variance within the population without creating a new source of reference bias from variants that are in other genomes that cannot be inherited. Furthermore, the risk of overfitting the pangenome to the population should be minimal due to all founders equally contributing to the progenies (Thompson et al., 2024). To understand the differences between a population-specific pangenome and a species-level

pangenome for use as a genotyping reference, we constructed a separate extended pangenome that includes genomes of three peanut genotypes, ‘Shitouqi’ (Zhuang et al. 2019), ‘Fuhuasheng’ (Chen et al. 2019) from China, and ‘Bailey II’ from North Carolina (Newman et al. 2023) along with the PeanutMAGIC founder genomes to serve as a species level pangenome representative. The extended pangenome has a total of 10,721,659 variants, adding 7,959,493 new variants (2.88 times of PeanutMAGIC pangenome total variants). These variants span throughout the genomes and are not strictly in defined regions that can be simply identified (Supplementary Figs. 3.2&3.3). The excessive variation in the extended pangenome that is not included in the population could increase genotyping errors if used for genotype calling, making species level pangenomes improper for mapping population applications. Therefore, the PeanutMAGIC pangenome can provide a comprehensive library of inherited haplotypes that is specific to the PeanutMAGIC population. This resource serves as an unbiased reference for accurate genotyping, trait mapping, and identification of candidate variations associated with phenotypes of interest.

PeanutMAGIC pangenome markers reveal chromosomal recombination patterns

The PeanutMAGIC population has 3,187 recombinant inbred lines (RILs) and the MAGIC Core is a subset of 310 RILs to represent the entire PeanutMAGIC population. The MAGIC Core subset has been sequenced using Illumina technology with an average coverage of 1x (Thompson et al. 2024). A total of 463,273 markers, including SNP, InDel, and complex variants (sites with more than two variants including InDels and multi-allelic SNPs) were identified for the MAGIC Core using three criteria: a variant is present in at least 25% of the MAGIC Core, in the PeanutMAGIC pangenome, and in at least one founder linear genome (Supplementary Fig. 3.4). Of the pangenomic markers, founder-specific markers preserve the

origin of a locus and enable accurate tracking of recombination points throughout the breeding scheme. A total of 413,488 unique markers (variant possessed by only one founder) were identified, including 336,941 SNPs, 36,268 InDels, and 4,697 complex sites (Fig. 3.2a). The high number of SNP markers, compared to InDel and complex markers, could be attributed to the limitations of low coverage sequencing for marker selection (Fig. 3.2b). In the case where structural variants lacked consistent and sufficient coverage, differentiating them across the population becomes challenging. Structural variants often contain repetitive sequences from neighboring regions, leading to ambiguous read mapping to both present and absent variants. This ambiguity renders structural variants less informative for marker selection. The number of founder-specific markers ranged from 5,023 in 97R to 241,971 in TNV (Fig. 3.2c). Notably, a majority of TNV specific markers (240,955; 99.58%) are located on Chr.09, which harbors the diploid *A. cardenasii* introgression (Holbrook et al., 2017). The founder C20 has 87,611 markers, highlighting its unique variants derived from a different subspecies and germplasm collection originating in China (Liang et al., 2005) (Fig. 3.2c).

The first 6.3 Mb of Chr.05 and Chr.15 lack markers due to the homeologous recombinant nature of these regions (Bertioli et al., 2019). To illustrate this, we aligned the founder TR to the diploid progenitors *A. duranensis* (AA) and *A. ipaensis* (BB) (Supplementary Fig. 3.5). The alleles within these locations can move, in blocks, between maternal Chr.05 and paternal Chr.05 in addition to maternal or paternal Chr.15. This complexity of tetrasomic inheritance makes it challenging to assign a single location for a marker site or to pass through Mendelian segregation filtering. To accurately determine where the marker resides and track recombination between Chr.05 and Chr.15 was challenging and difficult to describe with current genotyping methods, therefore, these regions were excluded from marker calling to not perpetuate error from

genotyping. This issue is consistent between the single-reference-based markers and pangenome-based markers and are not present in either marker set (Thompson et al., 2024).

To visualize recombination patterns within the PeanutMAGIC population, we analyzed r^2 values in a sliding window of 300 markers (Fig. 3.2d). The level of recombination varied both on the macro and micro scale across the genome. The chromosomes Chr.02, Chr.03, Chr.04, Chr.05, Chr.07, Chr.08, Chr.12, Chr.14, Chr.15, Chr.18, and Chr.20 exhibited consistent recombination levels across their respective structures ($r^2 < 0.5$), suggesting the detectable alleles on these chromosomes segregated normally. In contrast, Chr.01, Chr.06, Chr.09, Chr.10, Chr.11, Chr.13, Chr.16, Chr.17, and Chr.19 contain regions with elevated LD ($r^2 > 0.5$), indicating that alleles in these areas do not observe independent assortment (Fig 3.2d). The difference in recombination patterns may be the product of chromosome sequence divergence and structural variation facilitating regions that are more compatible than others, particularly from founders with increased genetic diversity (NC, C20, TNV).

Upon further examination, Chr.01, Chr.10, Chr.17, and Chr.19 exhibit large, distinct pericentromeric blocks with limited recombination, compared to the telomeric regions of the respective chromosomes. Chr.01 has a unique pericentromeric region derived from the founder NC (Supplementary Fig. 3.6). C20 is the donor for the unique pericentromeric regions of Chr.10 and Chr.17 (Supplementary Figs. 3.7 & 3.8). Interestingly, a different number of RILs possess the C20-based pericentromere on Chr.10 (27 RILs, 8.71%) compared to Chr.17 (54 RILs, 17.42%). This difference allows two factors to influence the r^2 values for the respective chromosome: the number of lines with the distinct pericentromere and the level of recombination within the pericentromeric region. Additionally, the unique pericentromere of Chr.19 originates from the founder 13M, a cultivar that possesses considerably less founder-specific alleles than

NC and C20 but maintains a unique pericentromere (Supplementary Fig. 3.9). These findings provide insights into recombination levels and inheritance patterns of pericentromeric regions of cultivated peanut chromosomes.

The chromosomes Chr.09, Chr.11, and Chr.13 displayed specific regions resistant to recombination (Supplementary Figs. 3.10-3.12). On Chr.09, the regions originated from the well-known *A. cardenasii* introgression, which spans the majority of Chr.09 on TNV. Notably, these regions recombine within the population as micro-fragments rather than large fragments, contrary from previous characterization in biparental populations (Clevenger et al 2017, Chu et al. 2016) and in PeanutMAGIC (Thompson et al. 2024), suggesting pangenome-based markers improve the ability to track such variants that have been previously undetectable (Supplementary Fig. 3.10). The elevated LD region on Chr.11 was derived from a block of alleles from the founder NC, suggesting reduced compatibility for recombination in that region compared to other NC alleles toward the middle and end of the chromosome, likely due to increased divergence in this block (Supplementary Fig. 3.11). Chr.13 contained three regions with elevated LD: one at the top, one at the bottom, and a central block. The top and bottom regions stemmed from C20 alleles that tended to segregate as a block. The central block harbored little marker diversity, suggesting that the elevated LD in this region is due to an inability to detect recombination rather than a true reduction in recombination (Supplementary Fig. 3.12).

The hindered ability to track recombination was also observed in homeologous chromosomes Chr.06 and Chr.16, where the pericentromeric regions showed limited marker variation. The lack of marker variation artificially inflated LD values due to the scarcity of markers able to trace recombination within these regions (Supplementary Fig. 3.13). The regions with low marker variance likely reflect existing variation that cannot be accurately tracked, possibly due to

increased repetitiveness or structural variants. In low coverage sequencing, such variants can result in the common allele being predominantly called in a block, creating areas with limited trackable variation, artificially inflating LD estimates (Supplementary Fig. 3.13). In contrast, regions of low genomic complexity such as those on Chr.13, provide limited information to accurately estimate recombination. This is due to an expanded physical window size between markers, which reduces the linkage signal and results in lower r^2 values (Supplementary Fig. 3.12). Such regions hinder genomic analyses and association studies due to their limited capacity to provide reliable recombination estimates. For regions with sufficient marker density and marker variation, recombination can be accurately tracked to identify areas of high LD or regions that segregate normally (e.g., Chr.11, Supplementary Fig. 3.11). Segregant founder-specific markers enable the ability to identify the origin of a particular genomic region, provided the region is distinct. Thus, a population-specific pangenome provides valuable insights to distinguish between regions of genuinely low complexity, inhibited recombination, and those merely lacking data.

A phylogenetic tree of the MAGIC Core shows an even distribution of founders and lines, with the exception of the TNV clade (Fig. 2e). This divergence stems from the inheritance and recombination of sizable *A. cardenasii* introgression blocks on Chr.09 from the donor TNV. Recombination of this introgression has been a point of interest to identify causal resistance genes for peanut root-knot nematode (RKN) (Clevenger et al., 2017; Chu et al., 2016; Chu et al., 2011). Genomic regions on Chr.09 have been previously identified with consistent high LD using single-reference-based markers (Thompson et al., 2024). However, with pangenome-based markers, variable LD patterns are observed in the introgressed region throughout the PeanutMAGIC population (Supplementary Fig. 3.14). This variability highlights the capability

of pangenome-based markers to capture recombination events with greater precision. By detecting micro-fragments of the introgression throughout the population, these markers provide insights into introgression inheritance and recombination processes that single-reference-based markers cannot resolve (Supplementary Figs. 3.10&3.14). Such enhanced resolution could be pivotal in the ongoing efforts to identify the elusive RKN resistance genes (Simpson 1991).

PeanutMAGIC pangenomic markers facilitate the identification of a third *AhFAD2* gene associated with high-oleic fatty acid content

An unbiased reference that integrates a comprehensive set of segregating variants improves marker retention and information within a population, thereby empowering association studies. To assess this hypothesis, peanut oleic acid phenotypes were selected due to their minimal environmentally influenced nature (Wang et al., 2024; Zhang et al., 2023), their status as the most studied yet unresolved peanut trait (Branch et al., 2022), and their importance for crop quality (Davis et al., 2016).

To understand the phenotypic distributions of oleic acid content in the MAGIC Core, seeds from the population were subjected to gas chromatography–mass spectrometry (GC-MS) for oleic acid quantification. Of the eight founders, four (13M, 97R, F07, TNV) exhibited high-oleic fatty acid phenotypes (>75%), one founder (WS16) displayed mid-oleic (55-75%), and three founders (C20, NC, TR) expressed low oleic content phenotypes (<55%). Within the MAGIC Core, a normal distribution was observed for low and mid-oleic content, with a distinct peak for high-oleic phenotypes (Supplementary Fig. 3.15). The normal distributions of low and mid-oleic content phenotypes suggest the involvement of more than two genes influencing oleic acid content, as reported by Branch et al. (2022).

A genome wide association study (GWAS) using PeanutMAGIC pangenome-based

markers identified two chromosomes, Chr.09 and Chr.19, associated with oleic acid content (Fig. 3a). A local association plot of Chr.19 pinpointed the known functional gene *AhFAD2B* at the correct location of 154.8 Mb ($P = 2.08x10^{-08}$) (Fig. 3.3b). Interestingly, four markers with higher significance values ($P = 6.42x10^{-09}$, $4.58x10^{-09}$, $3.06x10^{-09}$, $1.54x10^{-10}$) suggested the presence of another functional gene upstream of *AhFAD2B.* Using the TR protein annotations (Clevenger et al., 2016; Bertioli et al., 2019), we identified a third *AhFAD2* gene, 1.3 Mb upstream of *AhFAD2B*, neighboring the associated markers (Fig. 3.3c). This gene, named as *AhFAD2C*, lies in close proximity to *AhFAD2B*, making segregation challenging in small mapping populations, especially considering the presence of a third gene, *AhFAD2A*, on Chr.09. Additionally, the proximity of the two genes explained the increased association signal in this study and previous findings on Chr.19 compared to Chr.09 using biparental population (Pandey et al., 2014).

To compare with the pangenome-based study, we conducted GWAS using single-reference-based markers for oleic acid content. This analysis included a total of 138,151 markers, 3.35 times less than pangenome-based markers (Thompson et al., 2024). This study identified three chromosomes associated with oleic acid content: Chr.05, Chr.09, and Chr.19 (Fig. 3.3d). Notably, Chr.05 produced a likely false-positive signal, which was not detected with the PeanutMAGIC pangenome markers and had higher association values than known functional genes. The region identified on Chr.05 spans 3 Kb (90,216,715- 90,219,762) with no annotated genes in the region and no genes related to fatty acid biosynthesis were found near this region, suggesting that the signal is a false-positive (Supplementary Table 3.6). Furthermore, visualization of the PeanutMAGIC pangenome shows no variants in the region, although 30 single-reference markers were called (Supplementary Fig. 3.16). The identification of these markers in the region may stem from how single-reference markers are called, where the reads

are mapped to the reference and sequences that are similar can be interpreted as a polymorphism instead of mapping to a different location that is present in a different founder. These locations can be identified in the pangenome and are mapped directly. Additionally, the use of other association models with single-reference-based markers also fail to identify both Chr.09 and Chr.19, suggesting a genotyping error instead of a model issue (Supplementary Fig. 3.17). Upon examination of the local association plot of Chr.19 for single-reference-based markers, the identified signal was 1.1 Mb downstream (155.9 Mb) of *AhFAD2B* (154.8 Mb), demonstrating potential inaccuracies with single-reference-markers (Fig. 3.3e). The inaccurate signals on Chr.05 and Chr.19 undermine the legitimacy of the significant markers in the region containing *AhFAD2C*, thus could be overlooked in a single-reference study. These challenges exemplify some of the difficulties in understanding the genetic underpinnings of oleic acid content utilizing a single-reference framework and highlight the improved accuracy and precision of genomic associations when using a population-specific pangenome over a single-reference genome.

To validate these findings, we analyzed a biparental population derived from SunOleic 97R (97R) and NC94022 (NC), referred to as the 'S' population, which has been sequenced at 5x coverage (Agarwal et al. 2019, Khera et al. 2016). We constructed an 'S' population-specific pangenome and generated markers using the same approach to ensure that the differences between the populations were not due to methodological discrepancies. For the 'S' population, 113,056 population-specific pangenomic marker sites were identified for association with oleic acid content. Similar to the MAGIC Core study, two regions, Chr.09 and Chr.19, were associated with oleic acid content (Fig. 3.4a). A closer examination of Chr.19 revealed a genomic block associated with *AhFAD2B* and *AhFAD2C* (Fig. 3.4b). This region exhibited elevated LD ($r^2 >$ 0.3) in the 'S' population, suggesting that it generally segregates as a unit (Fig. 3.4c). In contrast,

the MAGIC Core displayed considerably lower LD ($r^2 < 0.2$) in this region to facilitate independent observation of *AhFAD2B* and *AhFAD2C* (Fig 3.4d). The findings of the biparental population suggest that only *AhFAD2A* and *AhFAD2B* may be required for the oleic acid phenotype, as minimal recombination occurred between *AhFAD2B* and *AhFAD2C* in the lines examined in this study. This finding highlights the difficulty of previous biparental population studies to resolve the genetic underpinnings of oleic acid and exemplifies the utility of multiparental populations for genomic studies.

The limited findings from biparental mapping populations can be exacerbated when applied as molecular markers in predictive breeding programs, with a greater number of recombinants. This issue was particularly evident in mid-oleic lines derived from a high-oleic population, which puzzled the research community (Branch et al. 2022). To illustrate this issue, we performed PCR-based genotyping on the PeanutMAGIC population to evaluate if tracking only *AhFAD2A* and *AhFAD2B* is sufficient for predicting oleic acid content, a common practice in marker-assisted selection (MAS) for peanut breeding (Chu et al., 2011). DNA was extracted from young leaf tissue of the MAGIC Core and genotyped using *AhFAD2A* and *AhFAD2B* PCR-based selectable markers previously employed in peanut breeding (Chu et al., 2011). Interestingly, two founders, TR and WS16, shared the same genotypes for both *AhFAD2A* and *AhFAD2B*, yet TR had low-oleic acid content while WS16 had mid-oleic acid content (Fig. 3.5a). The founder WS16 was found to possess a mutant genotype of *AhFAD2C*, whereas TR carried a wildtype genotype of *AhFAD2C.* Additionally, six lines were classified as mutants for both *AhFAD2A* and *AhFAD2B*, thus were identified as high-oleic lines using MAS but exhibited mid-oleic acid content phenotypes. The six lines would be excluded from future use in a breeding program due to the interpretation of contamination without information of *AhFAD2C*

showing that recombination occurred between mutant *AhFAD2B* and mutant *AhFAD2C* with a wildtype *AhFAD2C* (Fig. 3.5a). These findings indicate the necessity of including the third gene, *AhFAD2C*, in MAS application for peanut breeding to predict high-oleic phenotypes and to evaluate line purity.

We further examined *AhFAD2C* to identify potential causal variations. Visualization of the PeanutMAGIC pangenome revealed only two variations near *AhFAD2C*, one SNP located 907 bp upstream of the start codon and one 3-bp InDel situated 182 bp upstream of the SNP (Fig 3.5b). Sequence alignment to the founder linear genomes revealed four possible haplotypes for this region: C:A, C:ATTA, T:A, and T:ATTA (Supplementary Table 3.7). All high-oleic founders (13M, 97R, F07, TNV) and the mid-oleic founder (WS16) possess the C:ATTA haplotype. The low oleic founders exhibited distinct haplotypes: TR had the T:A haplotype, NC had the T:ATTA haplotype, and C20 had the C:A haplotype. To call variants for this specific location throughout the MAGIC Core, we utilized personalized pangenomes (Sirén et al., 2024) to accurately call the SNP and InDel site upstream of *AhFAD2C*. This approach allows for the utility of PeanutMAGIC pangenome subgraphs as a calling reference based on *k*-mer counts instead of the whole pangenome, reducing the potential for false variant calls. This yielded calls for the C/T SNP and the A/ATTA InDel variations upstream of *AhFAD2C* (Supplementary Fig. 3.18). We found this region to have parallels with other regions on Chr.19, making consistent calling for this site challenging and impractical in a single reference framework without a library of possible haplotypes to reduce ambiguous mapping and interpretation error (Supplementary Fig. 3.19). We performed an analysis of variance (ANOVA) on the three *AhFAD2* genes to determine their significance for oleic acid content. Results showed the Pr(>F) values of $<2.0 \times 10^{-16}$ for *AhFAD2A* and *AhFAD2B*, and 1.91×10^{-07} for *AhFAD2C*, indicating that all three genes

significantly affect oleic acid content in the MAGIC Core population (Supplementary Table 3.8). While *AhFAD2A* and *AhFAD2B* each have two alleles, *AhFAD2C* contains four haplotypes that can segregate within the PeanutMAGIC population.

To determine which variations had the greatest impact on oleic acid content, a Student's *t*-test was conducted among the different haplotypes (Fig. 5c). The largest single-site difference was observed between C:ATTA and T:ATTA, with a *p*-value of $4.6x10^{-05}$. Another notable difference was between C:ATTA and C:A, with a *p*-value of $1.3x10^{-03}$. These results suggest that the primary causal variant is the C/T SNP, as no significant change was observed with the addition of the InDel (from T:ATTA to T:A), and the SNP change (C:ATTA to T:ATTA) had a greater impact than the InDel change (C:ATTA to C:A). These analyses were repeated in the biparental 'S' population, yielding consistent results and confirming the significance of *AhFAD2C* for oleic acid content, with the C/T SNP identified as the primary variation upstream of *AhFAD2C* (Supplementary Table 3.9, Supplementary Fig. 3.20). Additionally, a post-hoc Tukey's honest significant difference (HSD) test was conducted for significant ANOVA factors in both the MAGIC Core and 'S' populations. We found that all significant ANOVA factors had significant differences in the MAGIC Core and 'S' populations, consistent with both the Student's *t*-test and ANOVA analyses (Supplementary Table 3.10&3.11). Of the *AhFAD2C* haplotype comparisons in the PeanutMAGIC, T:ATTA-C:ATTA, C:ATTA-C:A, and T:A-C:ATTA had significant differences (p adj < 0.05), consistent with the Student's *t*-test analysis. The ANOVA, Student's t-test, and post-hoc Tukey's HSD analyses of the MAGIC Core and 'S' populations support that *AhFAD2A*, *AhFAD2B* and *AhFAD2C* significantly contribute to oleic acid content, and the primary causal variation for *AhFAD2C* is the C/T SNP found in both the MAGIC Core and 'S' populations. These findings underscore the value of using a population-

specific pangenome to improve the understanding of the genetic basis of important traits.

We further explored the variance of *AhFAD2C* with low coverage, long read whole genome sequencing of 94 MAGIC Core lines to verify the genotypes of the low coverage, short read whole genome sequencing. The long-read data was aligned with the PeanutMAGIC pangenome to generate calls for *AhFAD2C*. Of the six lines that were called mutant for *AhFAD2A* and *AhFAD2B*, four lines had information of the region, and two lines failed to have coverage for *AhFAD2C* variants (Supplementary Table 3.12). One line, MG1404 was genotyped with the long read data as mutant for *AhFAD2A*, *AhFAD2B*, and *AhFAD2C*, suggesting that this genotype is high-oleic. Previously, MG1404 was genotyped as mutant for *AhFAD2A* and *AhFAD2B* with wildtype *AhFAD2C* and expressed as a mid-oleic phenotype (Fig. 3.5a). Because of the change in the *AhFAD2C* genotype, we performed high-throughput seed phenotyping to determine if the mid-oleic line MG1404 was contaminated with high-oleic seed. We tested 839 MG1404 seeds and identified 226 seeds (26.94%) were high-oleic, suggesting that this line is no longer a pure line (Supplementary Fig. 3.21). The impurity of MG1404 exemplifies the persistent challenges which peanut breeders and growers face to ensure consistency quality for confectionary products. This contamination issue would have been undetected using only *AhFAD2A* and *AhFAD2B* genotyping, demonstrating the necessity of *AhFAD2C* to predict high-oleic phenotypes and to evaluate purity.

To explore structural difference among the three *AhFAD2* genes, protein structure prediction was conducted using Alphafold (v2.3.1; Jumper et al., 2021). The predicted protein structures of the three *AhFAD2* genes exhibited similar α-helix motifs, a small β-sheet motif on the exterior and a long unstructured region spanning the first 50 amino acids (Fig. 3.5d). Although AhFAD2C had a shorter sequence than AhFAD2A and AhFAD2B, the predicted

structure retained the core structural features, suggesting functional relevance with reduced activity (Supplementary Fig. 3.22). The ANOVA significance of *AhFAD2C* was less than *AhFAD2A* and *AhFAD2B*, which agrees with AhFAD2C having less activity than AhFAD2A and AhFAD2B. These findings further support *AhFAD2C* as a functional gene influencing oleic fatty acid phenotypes in cultivated peanut.

Discussion

Segregating populations are essential resources for trait mapping and genetic improvement in breeding. However, two major challenges must be addressed to effectively utilize these populations: increased recombination events and the accuracy of genotyping calling. The PeanutMAGIC population addresses these issues through fine recombination of multiple genetic donors and a comprehensive library of inheritable haplotypes. We found that using the extended pangenome at the species-level with additional genomes considerably increased the number of non-representative variants, potentially leading to more marker calling errors. The population-specific pangenome resulted in high-quality genotyping calling and enabled accurate and precise association.

In this study we successfully used pangenome-based markers to track founder origin of chromosomal segments within a multiparent population. We also traced the founder origins for genomic regions that were resistant to recombination, offering unprecedented insight into the recombination of cultivated peanut. We highlighted the benefits of population-specific pangenome-based genotyping compared to single-reference genotyping for trait association studies. By relying on population-specific reference, we were able to reduce false signals made by single-reference association and offered increased resolution to identify *AhFAD2C*, near *AhFAD2B*, a crucial answer to the mystery of mid-oleic acid phenotypes in high-oleic breeding

populations, which puzzled the community. This was validated by using a biparental 'S' population and highlighted the difficulties of biparental population studies to resolve closely linked putative functional genes. Although the mechanism behind *AhFAD2C* is inferred, future studies can validate the functionality and role of *AhFAD2C* to facilitate high-oleic phenotypes using comparative long-read expression data, enzyme activity assays, and mutational studies. These findings highlight the power of a population-specific pangenome to capture genetic variation and recombination patterns that would be limited in a single-reference framework.

Methods

Plant Materials and Short Read Sequencing

The PeanutMAGIC population has been described in detail previously (Thompson et al., 2024). Briefly, eight founder parents: 'Georgia-13M' (13M), 'SunOleic 97R' (97R), 'GT-C20' (C20), 'Florida-07' (F07), 'NC94022' (NC), 'TifNV-High O/L' (TNV), 'Tifrunner' (TR), and 'GP-NC WS16' (WS16) were crossed in a simple funnel-like design. The critical 8-way cross consisted of 150 successful 4-way pairs, generating 950 unique F_1 offspring. The population was advanced to generate 3,187 $F_{2:7}$ RILs. A subset of 310 RILs were randomly selected for this study termed the MAGIC Core. These RILs were subject to CTAB DNA extraction followed by low-coverage (1x) Illumina sequencing.

The 'S' population has previously been described (Qin et al., 2012; Pandey et al., 2014; Agarwal et al., 2019). Briefly, this bioparental mapping population was derived from SunOleic 97R and NC94022 of 352 RILs. SunOleic 97R is a high-oleic fatty acid peanut cultivar. The two parents of the 'S' population are founders of the PeanutMAGIC. For this study, 144 RILs from the 'S' population were selected and sequenced at 5x with Illumina technology (Agarwal et al., 2019).

Founder Assembly and Pangenomes Construction

The DNA of seven founders of the PeanutMAGIC was extracted using a high molecular-weight DNA extraction and sequenced using PacBio Sequel II or Revio sequencing platforms at 25x coverage using the CCS mode to generate PacBio HiFi reads. The eighth founder Tifrunner (TR) (version 2 had previously been sequenced using PacBio and Illumina sequencing technologies (Bertioli et al., 2019). PacBio reads were assembled using hifiasm (v0.19.6) (Cheng et al., 2021) with the parameters '–u –l0'. The founders 13M, 97R, C20, and F07 were scaffolded with Hi-C data and polished using Juicer (Durand et al., 2016), Juicebox (Durand et al., 2016), and 3d-DNA (Dudchenko et al., 2017) workflow. The founders NC, TNV, and WS16 were scaffolded using RagTag (Alonge et al., 2022) with TR as the reference.

The resulting chromosome-level assemblies were aligned to TR as reference to assess quality and address erroneous assembly and orientation using minimap2 (Li, 2021) and dotPlotly for visualization. Assembly completeness was evaluated using BUSCO (v5.5.0) (Manni et al., 2021) with the 'fabales_odb10' database.

A population-specific pangenome for PeanutMAGIC was constructed using the Mini-Graph Cactus Pangenome Pipeline (v2.6.7) (Armstrong et al., 2020; Hickey et al., 2023) with TR as the reference genome. To generate a population-specific pangenome for the 'S' population, the 97R and NC assemblies were unified using the same pipeline, with NC as the reference. To assess the graphs vg (v1.56.0) (Garrison et al., 2018) subcommands 'stats' and 'deconstruct' (Liao et al., 2023) were used to extract components and variant locations for each pangenome.

Syntenic Block Analysis

Synteny information was analyzed using the SynMap tool on the CoGe platform (Lyons et al. 2008) using TR as reference. Only coding sequences were considered to identify unique

gene parallels between subgenomes, minimizing oversampling of low-complexity regions.

Khufu Pangenomic Marker Calling

KhufuPAN requires a graphical fragment assembly (GFA) file. The file should include the genomes forming the pangenome and a well-assembled reference genome (designated null reference), which is used to assign the coordinates of SNPs and structural variants. Minigraph-Cactus (https://github.com/ComparativeGenomicsToolkit/cactus) (Hickey et al., 2023) is recommended for creating the GFA file for the Khufu environment. The first step in the environment is KhufuPAN bootstrapping. The GFA is deconstructed to produce a parental VCF file. Then a series of filters is applied to remove odd alleles, low-quality variants, monomorphic variants, or those having polymorphism only with the null reference genome to generate a Filtered-Variant set. All files are packed into a single folder, which can be reused for multiple purposes.

KhufuVAR is a KhufuPAN sub-tool used to call and filter GFA variants for markers. Raw FASTQ files, short or long-reads, are applied to Khufu-core (https://www.hudsonalpha.org/khufudata/), and then mapped to the GFA file using vg giraffe (https://github.com/vgteam/vg/wiki/Mapping-short-reads-with-Giraffe) (Sirén et al., 2021). Calls under the minimum depth cutoff are masked, and variants overlapping with the Filtered-Variant set are extracted. Variants are segregated differently based on the population structure. Therefore, another series of filters are applied, i.e., minor allele frequency (>0.01), polymorphism, and the percentage of missing data (>75% missing). To efficiently utilize computational resources, KhufuVAR splits the process into two parts. Within the first part, samples run in batches, and every batch is aligned with the Filtered-Variant set and calculates measurements for the population filters. The second part combines the batches and runs the

population filters. The final output calls are exported in Khufu panmap format, which facilitates pangenome-based markers for genomic analysis (Supplementary Appendix). The generated panmap file can then be used to extract a rds file for Cyclops to visualize markers throughout the population in chromosomal units (https://w-korani.shinyapps.io/cyclops_eye_ii/).

Peanut MAGIC Core Pangenome-Based Genomic Characterization

To examine marker segregation within the MAGIC Core, r^2 values were calculated in a 300-marker sliding window on TASSEL 5.0 software (Bradbury et al., 2007) and plotted according to the physical location. The 300-marker window size was selected based on the increased marker density relative to the sliding window used in Thompson et al. (2024). To maintain comparable physical window sizes, the marker count per window was scaled up according to the average fold increase in markers (3x), simplifying comparison between single and pangenome references. For the 'S' population, r^2 values were calculated in a 100-marker sliding window to ensure equal physical distance window sizes.

The neighbor joining tree was constructed in R using the *ape* package (Paradis and Schliep, 2018). The workflow involved generating a distance matrix using 'dist.gene', converting the matrix into a tree with 'nj', and plotting the tree with 'plot', following the approach outlined in Thompson et al. (2024).

Oleic Fatty Acid Contents and GWAS Analysis

To collect phenotypic data for association studies, the seeds were planted in three replications in the field for two years along with the parental lines. The harvested seeds were analyzed chemically in USDA-ARS laboratory in Griffin, GA, for oil composition and content by analyzing five seeds per line using gas chromatography–mass spectrometry (GC-MS) (Pandey et. al., 2014).

PeanutMAGIC specific pangenome markers were used to calculate a distance matrix for multidimensional scaling (MDS) to control population structure. A general linear model was applied for GWAS analysis in TASSEL 5.0 software (Bradbury et al., 2007), using oleic acid contents and genotype data from three sources: MAGIC Core pangenome-based markers, MAGIC Core single-reference markers, and the 'S' pangenome markers. Each genotype source was paired with its respective phenotypes and MDS to maintain consistent parameters across associations.

PCR Genotyping

Perfect markers for *AhFAD2A* and *AhFAD2B* were previously developed and optimized for peanut breeding applications (Chu et al., 2011). Fresh leaf tissue was collected, and DNA was extracted using a NaOH solution and a TE buffer dilution. Diluted DNA was then used for marker for melting curve and KASP genotyping techniques, following the methods as described (Chu et al., 2011).

Personalized Pangenome Genotyping

Sequencing reads for peanut MAGIC Core were used to generate personalized pangenomes for each individual following the methods detailed in the vg GitHub wiki (https://github.com/vgteam/vg/wiki/Haplotype-Sampling) (Sirén et al., 2024). First, kmc was used to count *k*-mers from the sequence reads (Kokot et al., 2017). These *k*-mer counts were then incorporated into haplotype sampling with vg giraffe (Sirén et al., 2021). The subsequent GBZ and minimizer files were processed with vg 'pack' to compute read support for genotyping (Garrison et al., 2018). Finally, genotyping calls were made using vg 'call' (Hickey et al., 2020). This workflow generated personalized pangenomes and genotypes that were used to extract variation information for *AhFAD2C* across MAGIC Core and the 'S' populations.

Statistical Testing

All phenotypic statistical analyses were conducted and visualized using R statistical software (v4.4.1; R Core Team, 2024) with the *ggplot2* package (Wickham, 2016). ANOVA was performed with the base 'aov' function. The Student's *t*-test was conducted using the 'stat_compare_means(method = "t.test")' function from the *ggpubr* package (Kassambara, 2023). The post-hoc Tukey's HSD was performed with the base 'TukeyHSD' function.

Whole Genome Long Read Low Coverage Sequencing

High molecular weight DNA was extracted from 94 MAGIC Core lines using the unofficial Pacbio high-throughput gDNA workflow for plants (https://www.pacb.com/support/documentation/). The subsequent DNA was then multiplexed and formed libraries using Pacbio Hifi plex prep kit for Revio sequencing. The reads were then separated into their respective genotypes to generate raw long reads. The long reads were mapped to the PeanutMAGIC pangenome using the personalized pangenome genotyping approach and verified using minimap (Li, 2021).

In silico protein prediction

Protein structures of three *AhFAD2* genes were computationally predicted using Alphafold (v2.3.1) (Jumper et al., 2021). Annotated protein sequences were used to generate monomer structures for *AhFAD2A*, *AhFAD2B*, and *AhFAD2C*. The databases used for structure prediction included the BFD database "bfd_metaclust_clu_complete_id30_c90_final_seq.sorted_opt", the 'Mgnify' database from May 2022, the 'pdb70' database from March 2022, the 'Uniclust 30' database from March 2021, and the 'Uniref 30' database from March 2021.

References

Agarwal, G. et al. A recombination bin-map identified a major QTL for resistance to tomato spotted wilt virus in peanut (Arachis hypogaea). Scientific Reports 9, (2019).

Alonge, M. et al. Automated Assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biology 23, (2022).

Alper, C. M., & Mattes, R. D. (2003). Peanut consumption improves indices of cardiovascular disease risk in healthy adults. *Journal of the American College of Nutrition*, *22*(2), 133–141. https://doi.org/10.1080/07315724.2003.10719286

Armstrong, J. et al. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. Nature 587, 246–251 (2020).

Bertioli, David J, Ozias-Akins, P., Chu, Y., Dantas, K. M., Santos, S. P., Gouvea, E., Guimarães, P. M., Leal-Bertioli, S. C., Knapp, S. J., & Moretzsohn, M. C. (2014). The use of SNP markers for linkage mapping in diploid and tetraploid peanuts. *G3 Genes|Genomes|Genetics*, *4*(1), 89–96. https://doi.org/10.1534/g3.113.007617

Bertioli, David J., Jenkins, J., Clevenger, J., Dudchenko, O., Gao, D., Seijo, G., Leal-Bertioli, S. C., Ren, L., Farmer, A. D., Pandey, M. K., Samoluk, S. S., Abernathy, B., Agarwal, G., Ballén-Taborda, C., Cameron, C., Campbell, J., Chavarro, C., Chitikineni, A., Chu, Y., … Schmutz, J. (2019). The genome sequence of segmental allotetraploid peanut arachis hypogaea. *Nature Genetics*, *51*(5), 877–884. https://doi.org/10.1038/s41588-019-0405-z

Bertioli, David John, Cannon, S. B., Froenicke, L., Huang, G., Farmer, A. D., Cannon, E. K., Liu, X., Gao, D., Clevenger, J., Dash, S., Ren, L., Moretzsohn, M. C., Shirasawa, K.,

Huang, W., Vidigal, B., Abernathy, B., Chu, Y., Niederhuth, C. E., Umale, P., … Ozias-Akins, P. (2016). The genome sequences of arachis duranensis and Arachis ipaensis, the diploid ancestors of cultivated peanut. *Nature Genetics*, *48*(4), 438–446. https://doi.org/10.1038/ng.3517

Bradbury, P. J. et al. Tassel: Software for association mapping of complex traits in diverse samples. Bioinformatics 23, 2633–2635 (2007).

Branch, W. D., Brown, N. & Perrera, M. A. Inheritance of mid-oleic fatty acid ratio seed trait in peanut. Peanut Science 49, 26–29 (2022).

Branch, W. D. (2014). Registration of 'Georgia-13M' peanut. *Journal of Plant Registrations*, *8*(3), 253–256. https://doi.org/10.3198/jpr2013.11.0071crc

Burow, M. D., Simpson, C. E., Starr, J. L., & Paterson, A. H. (2001). Transmission genetics of chromatin from a synthetic amphidiploid to cultivated peanut (*arachis hypogaea*L.): Broadening the gene pool of a monophyletic polyploid species. *Genetics*, *159*(2), 823–837. https://doi.org/10.1093/genetics/159.2.823

Chen, X. et al. Sequencing of cultivated peanut, Arachis hypogaea, yields insights into genome evolution and oil improvement. Molecular Plant 12, 920–934 (2019).

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nature Methods 18, 170–175 (2021).

Chu, Y., Ramos, L., Holbrook, C. C., & Ozias-Akins, P. (2007). Frequency of a loss-of-function mutation in oleoyl-pc desaturase (*ahfad2a*) in the mini-core of the U.S. peanut germplasm collection. *Crop Science, 47*(6), 2372–2378. https://doi.org/10.2135/cropsci2007.02.0117

Chu, Ye, 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. et al. Marker-assisted selection to pyramid nematode resistance and the high oleic trait in peanut. The Plant Genome 4, 110–117 (2011).

Chu, Y. et al. Identification of rare recombinants leads to tightly linked markers for nematode resistance in peanut. Peanut Science 43, 88–93 (2016).

Clevenger, J. P., & Ozias-Akins, P. (2015). Sweep: A tool for filtering high-quality snps in polyploid crops. *G3 Genes|Genomes|Genetics, 5*(9), 1797–1803. https://doi.org/10.1534/g3.115.019703

Clevenger, J. P., Korani, W., Ozias-Akins, P., & Jackson, S. (2018). Haplotype-based genotyping in Polyploids. *Frontiers in Plant Science, 9*. https://doi.org/10.3389/fpls.2018.00564

Clevenger, J., Chu, Y., Scheffler, B., & Ozias-Akins, P. (2016). A developmental transcriptome map for allotetraploid Arachis hypogaea. *Frontiers in Plant Science, 7*. https://doi.org/10.3389/fpls.2016.01446

Clevenger, J. et al. Gene expression profiling describes the genetic regulation of Meloidogyne arenaria resistance in arachis hypogaea and reveals a candidate gene for resistance. Scientific Reports 7, (2017).

Culbreath, A. K., Gorbet, D. W., Martinez-Ochoa, N., Holbrook, C. C., Todd, J. W., Isleib, T. G., & Tillman, B. (2005). High levels of field resistance to tomato spotted wilt virus in peanut breeding lines derived from Hypogaea and hirsuta botanical varieties. *Peanut Science*, *32*(1), 20–24. https://doi.org/10.3146/0095-3679(2005)32[20:hlofrt]2.0.co;2

Davis, J. P., Price, K. M., Dean, L. L., Sweigart, D. S., Cottonaro, J. M., & Sanders, T. H. (2016). Peanut oil stability and physical properties across a range of industrially relevant oleic acid/linoleic acid ratios. *Peanut Science*, *43*(1), 1–11. https://doi.org/10.3146/0095-3679-43.1.1

Danilevicz, M. F., Tay Fernandez, C. G., Marsh, J. I., Bayer, P. E. & Edwards, D. (2020). Plant pangenomics: approaches, applications and advancements. *Curr. Opin. Plant Biol.* 54, 18–25. https://doi.org/10.1016/j.pbi.2019.12.005

Dudchenko, O. et al. De Novo Assembly of the Aedes aegypti genome using HI-C yields chromosome-length scaffolds. Science 356, 92–95 (2017).

Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution HI-C experiments. Cell Systems 3, 95–98 (2016).

Durand, N. C. et al. Juicebox provides a visualization system for HI-C contact maps with unlimited zoom. Cell Systems 3, 99–101 (2016).

Garrison, E. et al. Variation Graph Toolkit improves read mapping by representing genetic variation in the reference. Nature Biotechnology 36, 875–879 (2018).

Gorbet, D. W., & Knauft, D. A. (2000). Registration of 'SUNOLEIC 97R' peanut. *Crop Science*, *40*(4), 1190–1191. https://doi.org/10.2135/cropsci2000.0032rcv

Gorbet, D. W., & Tillman, B. L. (2009). Registration of 'florida-07' peanut. *Journal of Plant Registrations*, *3*(1), 14–18. https://doi.org/10.3198/jpr2008.05.0276crc

Guo, B., Khera, P., Wang, H., Peng, Z., Sudini, H., Wang, X., Osiru, M., Chen, J., Vadez, V., Yuan, M., Wang, C. T., Zhang, X., Waliyar, F., Wang, J., & Varshney, R. K. (2016). Annotation of trait loci on integrated genetic maps of arachis species. *Peanuts*, 163–207. https://doi.org/10.1016/b978-1-63067-038-2.00006-x

Guo, L. et al. (2025). Super pangenome of *Vitis* empowers identification of downy mildew resistance genes for grapevine improvement. *Nature Genetics*, 57, 741-753. https://doi.org/10.1038/s41588-025-02111-7

Herselman, L., Thwaites, R., Kimmins, F. M., Courtois, B., van der Merwe, P. J., & 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

Hickey, G. et al. Pangenome graph construction from genome alignments with Minigraph-Cactus. Nature Biotechnology 42, 663–673 (2023).

Hickey, G., Heller, D., Monlong, J., Sibbesen, J. A., Sirén, J., Eizenga, J., Dawson, E. T., Garrison, E., Novak, A. M., & Paten, B. (2020). Genotyping structural variants in pangenome graphs using the VG toolkit. *Genome Biology*, *21*(1). https://doi.org/10.1186/s13059-020-1941-7

Holbrook, C. C., Ozias-Akins, P., Chu, Y., Culbreath, A. K., Kvien, C. K., & Brenneman, T. B. (2017). Registration of 'tifnv-high O/L' peanut. *Journal of Plant Registrations*, *11*(3), 228–230. https://doi.org/10.3198/jpr2016.10.0059crc

Hou, Y., Gan, J., Fan, Z., Sun, L., Garg, V., Wang, Y., Li, S., Bao, P., Cao, B., Varshney, R. K., & Zhao, H. (2024). Haplotype-based Pangenomes reveal genetic variations and climate adaptations in moso bamboo populations. *Nature Communications*, *15*(1). https://doi.org/10.1038/s41467-024-52376-5

Isobe, S., Shirasawa, K., & Hirakawa, H. (2019). Current status in whole genome sequencing and analysis of Ipomoea spp.. *Plant Cell Reports*, *38*(11), 1365–1371. https://doi.org/10.1007/s00299-019-02464-4

Jones, J. B., Provost, M., Keaver, L., Breen, C., Ludy, M.-J., & Mattes, R. D. (2014). A randomized trial on the effects of flavorings on the health benefits of daily peanut consumption. *The American Journal of Clinical Nutrition*, *99*(3), 490–496. https://doi.org/10.3945/ajcn.113.069401

Jung, S., Powell, G., Moore, K., & Abbott, A. (2000). The high oleate trait in the cultivated peanut [arachis hypogaea L.]. II. molecular basis and genetics of the trait. *Molecular and General Genetics MGG*, *263*(5), 806–811. https://doi.org/10.1007/s004380000243

Jumper, J. et al. Highly accurate protein structure prediction with alphafold. Nature 596, 583–589 (2021).

Kassambara A., ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.6.0, https://rpkgs.datanovia.com/ggpubr/ (2023).

Khan, A. W. et al. Cicer super-pangenome provides insights into species evolution and agronomic trait loci for crop improvement in Chickpea. Nature Genetics 56, 1225–1234 (2024).

Kokot, M., Długosz, M., & Deorowicz, S. (2017). KMC 3: Counting and manipulating *k*-mer statistics. *Bioinformatics*, *33*(17), 2759–2761. https://doi.org/10.1093/bioinformatics/btx304

Korani, W., O'Connor, D., Chu, Y., Chavarro, C., Ballen, C., Guo, B., Ozias-Akins, P., Wright, G., & Clevenger, J. (2021). De Novo QTL-seq identifies loci linked to blanchability in peanut (arachis hypogaea) and refines previously identified QTL with low coverage sequence. *Agronomy*, *11*(11), 2201. https://doi.org/10.3390/agronomy11112201

Kris-Etherton, P. M., Hu, F. B., Ros, E., & Sabaté, J. (2008). The role of tree nuts and peanuts in the prevention of coronary heart disease: Multiple potential mechanisms. *The Journal of Nutrition*, *138*(9). https://doi.org/10.1093/jn/138.9.1746s

Li, H. (2021). New strategies to improve MINIMAP2 alignment accuracy. *Bioinformatics*, *37*(23), 4572–4574. https://doi.org/10.1093/bioinformatics/btab705

Li, X. et al. Large-scale gene expression alterations introduced by structural variation drive morphotype diversification in Brassica oleracea. Nature Genetics 56, 517–529 (2024).

Liu, Z., Wang, N., Su, Y., Long, Q., Peng, Y., Shangguan, L., Zhang, F., Cao, S., Wang, X., Ge, M., Xe, H., Ma, Z., Liu W., Xu, X., Li, C., Cao, X., Ahmad, X., Su, X., Liu, Y., Huang, G., Du, M., Liu, Z., Gan, Y., Sun, L., Fan, X., Zhang, C., Zhong H., Leng, X., Ren, Y., Dong, T., Pei, D., Wu, X., Jin, Z., Wang, Y., Liu, C., Chen, J., Gaut, B., Huang, S., Fang, J., Xiao, H., & Zhou, Y. (2024). Grapevine pangenome facilitates trait genetics and genomic breeding. *Nature Genetics* 56, 2804–2814 https://doi.org/10.1038/s41588-024-01967-5

Liang, X. Q., Holbrook, C. C., Lynch, R. E., & Guo, B. Z. (2005). B-1,3-glucanase activity in peanut seed (*arachis hypogaea*) is induced by inoculation with *aspergillus flavus* and copurifies with a conglutin-like protein. *Phytopathology®*, *95*(5), 506–511. https://doi.org/10.1094/phyto-95-0506

Liao, W.-W. et al. A draft human pangenome reference. Nature 617, 312–324 (2023).

López, Y., Nadaf, H. L., Smith, O. D., Connell, J. P., Reddy, A. S., & Fritz, A. K. (2000). Isolation and characterization of the Δ12-fatty acid desaturase in peanut (arachis hypogaea L.) and search for polymorphisms for the high oleate trait in Spanish market-type lines. *Theoretical and Applied Genetics*, *101*(7), 1131–1138. https://doi.org/10.1007/s001220051589

Lyons, E. et al. Finding and comparing syntenic regions among Arabidopsis and the Outgroups Papaya, Poplar, and Grape: CoGe with Rosids. Plant Physiology 148, 1772–1781 (2008).

Manni, M. et al. BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Molecular Biology and Evolution 38, 4647–4654 (2021).

Newman, C. S. et al. Initiation of genomics-assisted breeding in Virginia-type peanuts through the generation of a de novo reference genome and informative markers. Frontiers in Plant Science 13, (2023).

Pandey, M. K., Agarwal, G., Kale, S. M., Clevenger, J., Nayak, S. N., Sriswathi, M., Chitikineni, A., Chavarro, C., Chen, X., Upadhyaya, H. D., Vishwakarma, M. K., Leal-Bertioli, S., Liang, X., Bertioli, D. J., Guo, B., Jackson, S. A., Ozias-Akins, P., & Varshney, R. K. Development and evaluation of a high density genotyping 'Axiom_*Arachis*' array with 58K SNPs for accelerating genetics and breeding in groundnut (*Arachis* species). Scientific Reports, 7:40577. doi: 10.1038/srep40577, (2017).

Pandey, M.K., Wang, M.L., Qiao, L., Feng, S., Khera, P., Wang, H., Tonnis, B., Barkley, N.A., Wang, J., Holbrook, C.C., Culbreath, A.K., Varshney, R.K. and Guo, B.Z. 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. 15:133. doi: 10.1186/s12863-014-0133-4. (2014).

Paradis, E., & Schliep, K. APE 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics, 35(3), 526–528 (2018). with 58 K snps for accelerating genetics and breeding in Groundnut. *Scientific Reports*, *7*(1). https://doi.org/10.1038/srep40577 R Core Team, R: A language and environment for

statistical computing. R Foundation for Statistical computing, Vienna, Austria. https://www.R-project.org/ (2024).

Qin, H. et al. An integrated genetic linkage map of cultivated peanut (*Arachis hypogaea* L.) constructed from two RIL populations. *Theor. Appl. Genet*. 124, 653–664 (2012).

Scott, M. F. et al. Multi-parent populations in crops: A toolbox integrating genomics and genetic mapping with breeding. Heredity 125, 396–416 (2020).

Simpson, C. E. 1991. Pathways for introgression of pest resistance into *Arachis hypogaea* L. *Peanut Sci* 18:22-6.

Sirén, J., Eskandar, P., Ungaro, M. T., Hickey, G., Eizenga, J. M., Novak, A. M., Chang, X., Chang, P.-C., Kolmogorov, M., Carroll, A., Monlong, J., & Paten, B. (2024). Personalized pangenome references. *Nature Methods*. https://doi.org/10.1038/s41592-024-02407-2

Sirén, J., Monlong, J., Chang, X., Novak, A. M., Eizenga, J. M., Markello, C., Sibbesen, J. A., Hickey, G., Chang, P.-C., Carroll, A., Gupta, N., Gabriel, S., Blackwell, T. W., Ratan, A., Taylor, K. D., Rich, S. S., Rotter, J. I., Haussler, D., Garrison, E., & Paten, B. (2021). Pangenomics enables genotyping of known structural variants in 5202 diverse genomes. *Science*, *374*(6574). https://doi.org/10.1126/science.abg8871

Tallury, S. P., Isleib, T. G., Copeland, S. C., Rosas-Anderson, P., Balota, M., Singh, D., & Stalker, H. T. (2013). Registration of two multiple disease-resistant peanut germplasm lines derived from *arachis cardenasii* krapov. & W.C. Gregory, GKP 10017. *Journal of Plant Registrations*, *8*(1), 86–89. https://doi.org/10.3198/jpr2013.04.0017crg

Thompson E. et al. Genetic and genomic characterization of a multiparent advanced generation inter-cross (MAGIC) population of peanut (*Arachis hypogaea* L.). Crop Science, 2025

Varshney, R. K., Bertioli, D. J., Moretzsohn, M. C., Vadez, V., Krishnamurthy, L., Aruna, R., Nigam, S. N., Moss, B. J., Seetha, K., Ravi, K., He, G., Knapp, S. J., & Hoisington, D. A. (2008). The first SSR-based genetic linkage map for cultivated groundnut (arachis hypogaea L.). *Theoretical and Applied Genetics*, *118*(4), 729–739. https://doi.org/10.1007/s00122-008-0933-x

Vaughn, J. N., Branham, S. E., Abernathy, B., Hulse-Kemp, A. M., Rivers, A. R., Levi, A., & Wechter, W. P. (2022). Graph-based Pangenomics maximizes genotyping density and reveals structural impacts on fungal resistance in melon. *Nature Communications, 13*(1). https://doi.org/10.1038/s41467-022-35621-7

Wan, L., Li, B., Pandey, M. K., Wu, Y., Lei, Y., Yan, L., Dai, X., Jiang, H., Zhang, J., Wei, G., Varshney, R. K., & Liao, B. (2016). Transcriptome analysis of a new peanut seed coat mutant for the physiological regulatory mechanism involved in seed coat cracking and pigmentation. *Frontiers in Plant Science, 7*. https://doi.org/10.3389/fpls.2016.01491

Wang, H., Lei, Y., Wan, L., Yan, L., Lv, J., Dai, X., Ren, X., Guo, W., Jiang, H., & Liao, B. (2016). Comparative transcript profiling of resistant and susceptible peanut post-harvest seeds in response to aflatoxin production by aspergillus flavus. *BMC Plant Biology*, *16*(1). https://doi.org/10.1186/s12870-016-0738-z

Wang, M. L., Tonnis, B., Li, X., Benke, R., Huang, E., Tallury, S., Puppala, N., Peng, Z., & Wang, J. (2024). Genotype, environment, and their interaction effects on peanut seed

protein, oil, and fatty acid content variability. *Agronomy Journal*, *116*(3), 1440–1454. https://doi.org/10.1002/agj2.21559

Wickham, H. Create elegant data visualizations using the grammar of graphics. Create Elegant Data Visualizations Using the Grammar of Graphics. https://ggplot2.tidyverse.org/index.html, (2016).

Wu, S., Lau, K. H., Cao, Q., Hamilton, J. P., Sun, H., Zhou, C., Eserman, L., Gemenet, D. C., Olukolu, B. A., Wang, H., Crisovan, E., Godden, G. T., Jiao, C., Wang, X., Kitavi, M., Manrique-Carpintero, N., Vaillancourt, B., Wiegert-Rininger, K., Yang, X., … Fei, Z. (2018). Genome sequences of two diploid wild relatives of cultivated sweetpotato reveal targets for genetic improvement. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-06983-8

Wu, S., Sun, H., Hamilton, J. P., Mollinari, M., Gesteira, G. D., Kitavi, M., Yan, M., Wang, H., Yang, J., Yencho, G. C., Buell, C. R., & Fei, Z. (2024). *Phased Chromosome-Level Genome Assembly Provides Insight into the Origin of Hexaploid Sweetpotato*. https://doi.org/10.1101/2024.08.17.608395

Yencho, G. C., Pecota, K. V., Schultheis, J. R., VanEsbroeck, Z.-P., Holmes, G. J., Little, B. E., Thornton, A. C., & Truong, V.-D. (2008). 'covington' sweetpotato. *HortScience*, *43*(6), 1911–1914. https://doi.org/10.21273/hortsci.43.6.1911

Zhang, H., Yu, Y., Wang, M., Dang, P., & Chen, C. (2023). Effect of genotype-by-environment interaction on oil and oleic fatty acid contents of cultivated peanuts. *Horticulturae*, *9*(12), 1272. https://doi.org/10.3390/horticulturae9121272

Zhang, Y. et al. Telomere-to-telomere Citrullus super-pangenome provides direction for watermelon breeding. Nature Genetics 56, 1750–1761 (2024).

Zhuang, W. et al. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. Nature Genetics 51, 865–876 (2019).

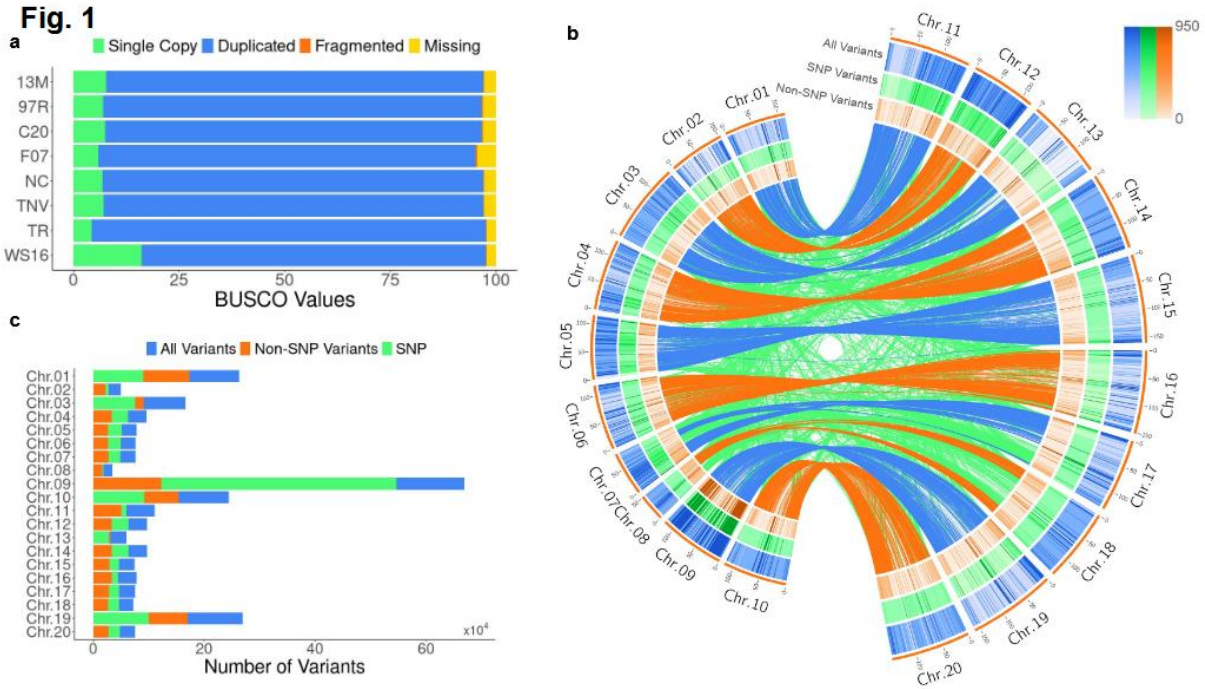


Fig. 3.1: **PeanutMAGIC pangenomic variation. a**, BUSCO values of the eight PeanutMAGIC founder assemblies, ‘Georgia-13M’ (13M), ‘SunOleic 97R’ (97R), ‘GT-C20’ (C20), ‘Florida-07’ (F07), ‘NC94022’ (NC), ‘TifNV-High O/L’ (TNV), ‘Tifrunner’ (TR), and ‘GP-NC WS16’ (WS16), showing high duplicated BUSCO values from similar subgenomes. **b**, Circos plot representing A (Chr.01-Chr.10, left side) and B (Chr.11-Chr.20, right side) subgenomes of cultivated peanut with outer orange bars as scales in Mb. The distribution of all variant sites within the PeanutMAGIC pangenome is represented as a blue heatmap (per 1 Mb; 0-950 sites), SNP variant locations are represented as a green heatmap (per 1 Mb; 0-950 sites), and all non-SNP variants are represented as an orange heatmap (per 1 Mb; 0-950 sites). Blue and orange lines represent homeologous syntenic blocks between A and B subgenomes and green lines represent non-homeologous syntenic blocks. **c**, Number of PeanutMAGIC pangenome variants per chromosome. Blue bars

represent all variants, orange bars represent non-SNP variants, and green bars represent SNP variants. Note: parts of the longer bars are overshadowed by smaller bars.

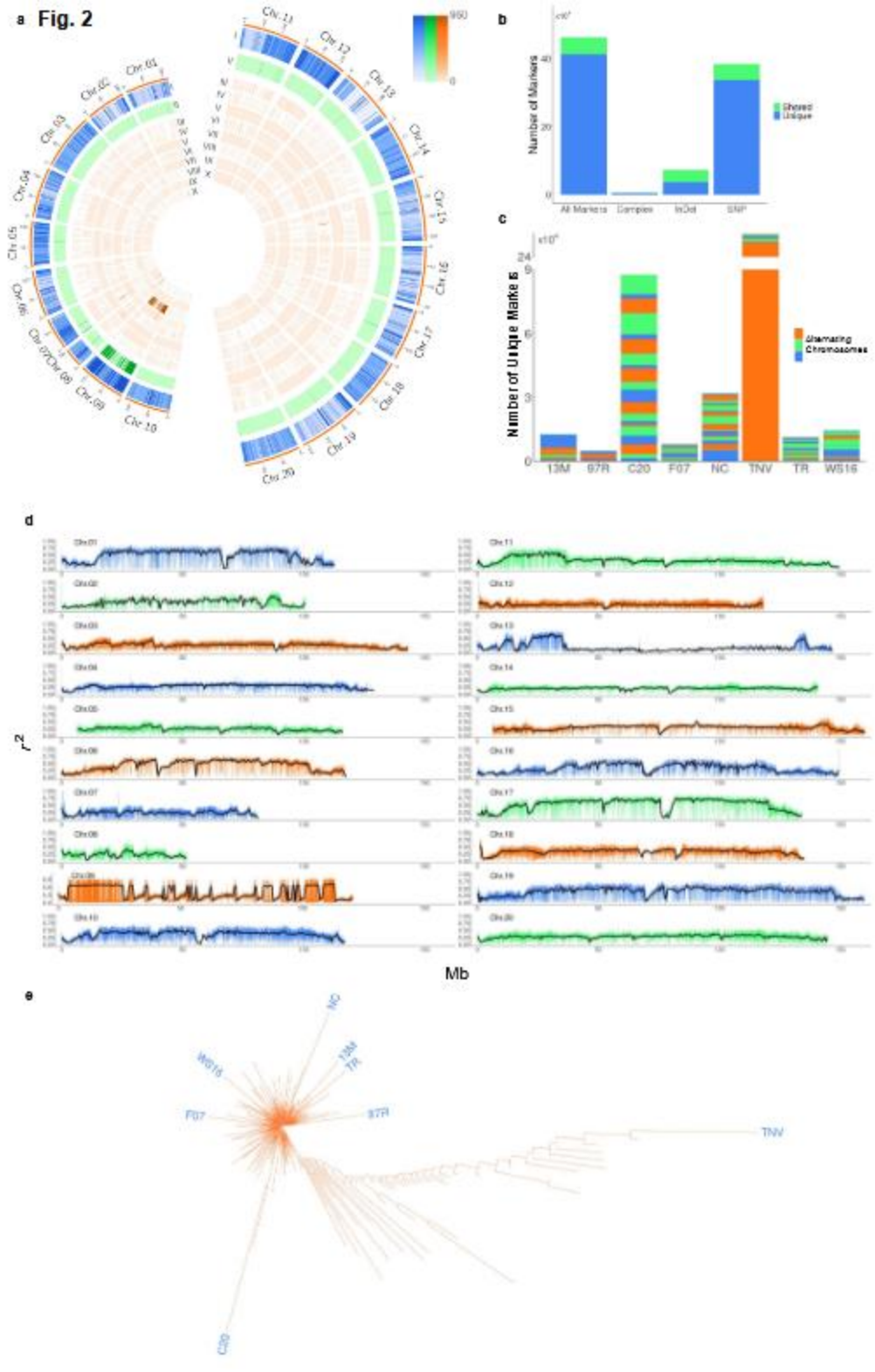

a Fig. 2
b
c
d
e
Number of Markers
Number of Unique Markers
Shared
Unique
Shared Chromosomes
All Markers
Complex
InDel
SNP
13M
97R
C20
F07
NC
TNV
TR
WS16
Mb
r^2

Fig. 3.2: **PeanutMAGIC pangenomic markers. a**, Circos plot representing A (Chr.01-Chr.10, left side) and B (Chr.11-Chr.20, right side) subgenomes of cultivated peanut with outer orange bars representing scales in Mb. I, Heatmap of all variant sites within the PeanutMAGIC pangenome, per 1 Mb. II, Heatmap of all pangenome markers retained for analysis, per 1 Mb. Heat map of founder-specific markers: 'Georgia-13M' (13M, III), 'SunOleic 97R' (97R, IV), 'GT-C20' (C20, V), 'Florida-07' (F07, VI), 'NC94022' (NC, VII), 'TifNV-High O/L' (TNV, VIII), 'Tifrunner' (TR, XI), 'GP-NC WS16' (WS16, X), per 1 Mb. **b**, The number of all markers, complex markers, InDel markers, and SNP markers. Blue represents the number of markers unique to a single founder and green represents the number of markers that are not exclusive to one founder as a percentage of the total bar. **c**, The number of markers unique to each founder. Alternating color bands represent the number of markers per chromosome from Chr.01 (bottom) to Chr.20 (top). **d**, r^2 values across chromosomes of the PeanutMAGIC Core population. Values were calculated through a sliding window of 300 markers. Individual markers are represented by blue, green or orange lines. The black line represents 0.5 Mb averages of r^2 values. **e**, Neighbor Joining phylogenetic tree of the PeanutMAGIC Core using PeanutMAGIC pangenomic markers. Founders are highlighted in blue.

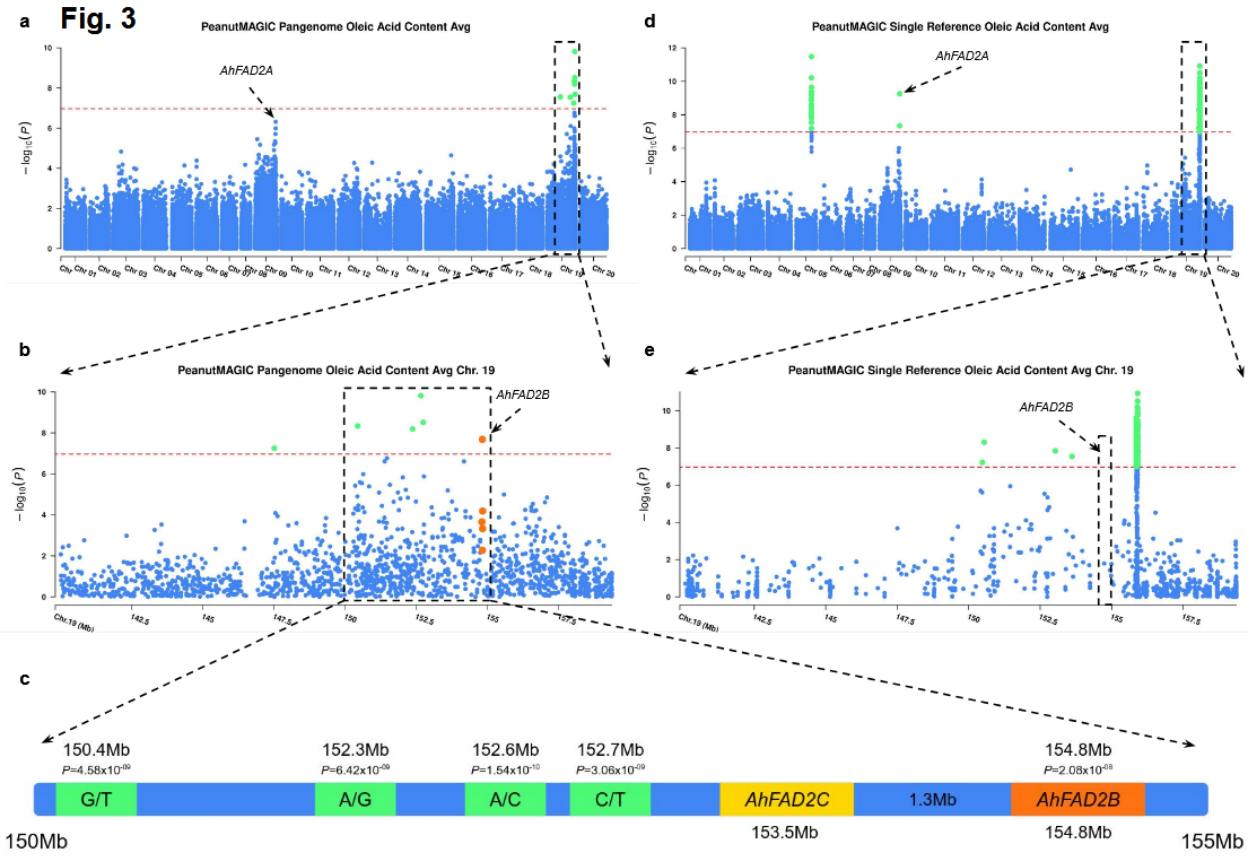


Fig. 3.3: **Oleic acid content association with PeanutMAGIC pangenomic and single reference markers. a**, Manhattan plot showing the association signals of pangenome-based markers with oleic acid content. A Bonferroni-corrected *P* value of 0.05 was used as a significant threshold ($P = 1.1x10^{-7}$), represented by a horizontal dashed red line. **b**, Local association plot of Chr.19 showing multiple significant locations. The orange highlight represents the location of the functional gene *AhFAD2B*. Regions upstream of a known functional gene have higher association values. **c**, Genetic view of the associated region showing significant markers and genes of interest. Markers are highlighted in green and denoted above the genome bar with *P* values and physical locations. The known functional gene *AhFAD2B* is highlighted orange. Between the associated markers and a known functional gene, a third *FAD2* gene was identified and termed *AhFAD2C*, highlighted in

yellow. **d**, Manhattan plot showing PeanutMAGIC Core single-reference marker signals associated with oleic acid content with an additional signal identified on Chr.05 that has not been reported in previous studies. **e**, Local association plot on Chr.19 showing a strong association downstream of the known *AhFAD2B* location, with no signal observed at the physical position of *AhFAD2B*.

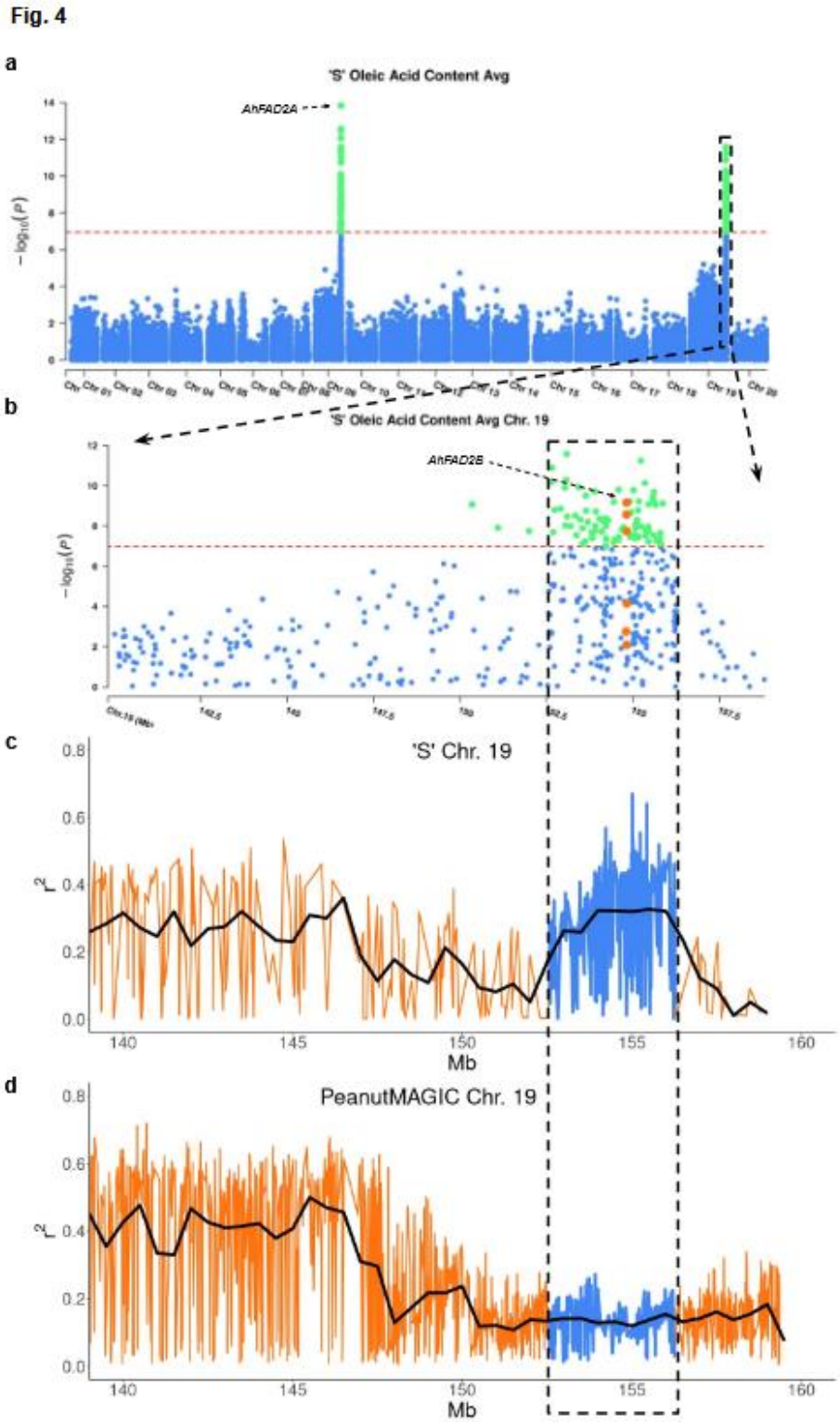


Fig. 3.4: **The 'S' biparental population (Qin et al., 2012, Pandey et al., 2014) pangenome oleic acid content association. a**, Manhattan plot showing 'S' pangenome marker signals

associated with oleic acid content. A Bonferroni-corrected *P* value of 0.05 was used as a significant threshold ($P = 1.1x10^{-7}$), represented by a horizontal dashed red line. **b**, Local association plot on Chr.19 showing a consistent associated region. The orange highlight represents the location of a functional gene *AhFAD2B*. **c**, The 'S' pangenome marker r^2 values of the associated region, indicating increased linkage disequilibrium. **d**, PeanutMAGIC pangenome marker r^2 values of the associated region, indicating lower linkage disequilibrium.

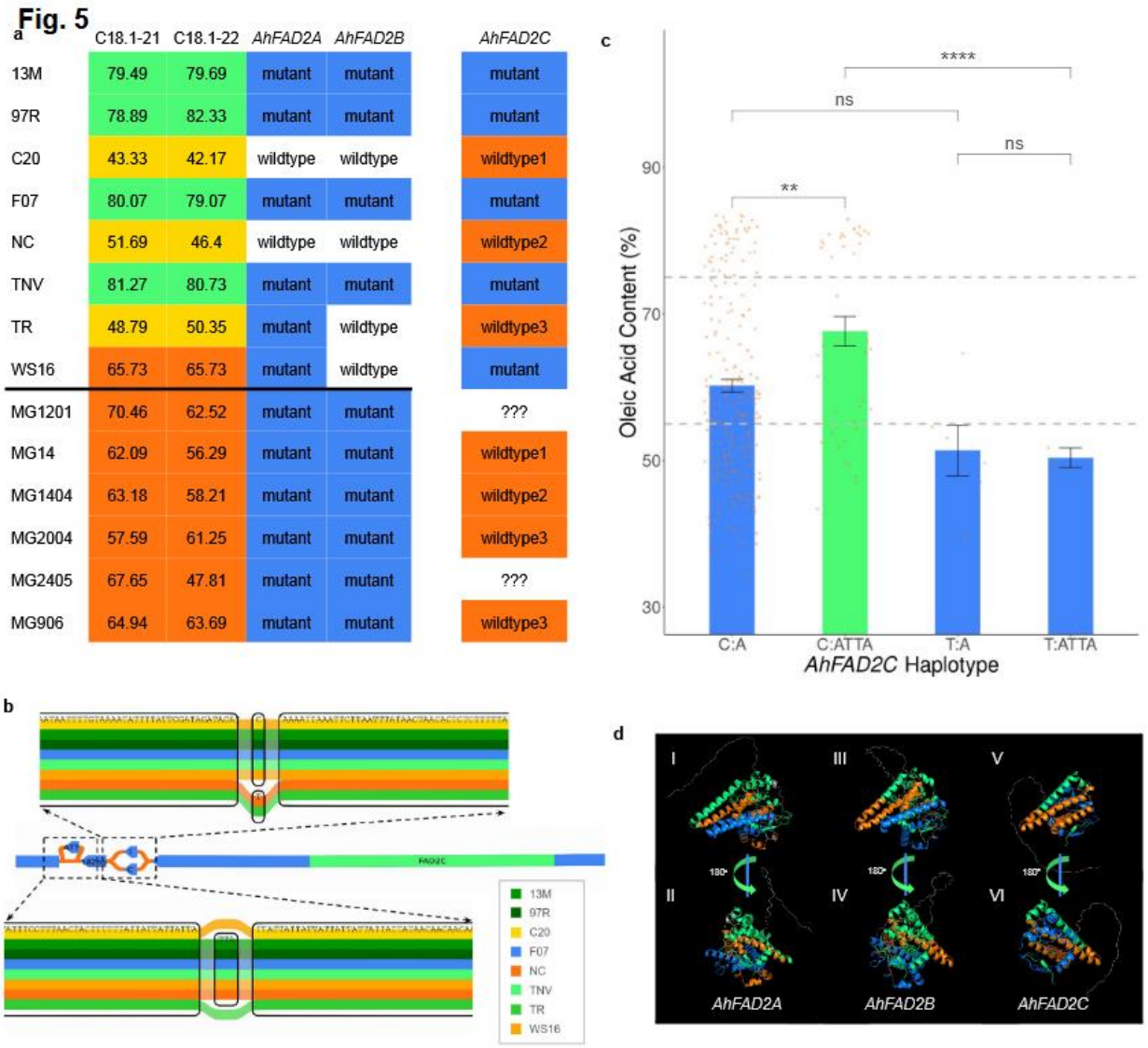


Fig. 3.5: **Functional variation of *AhFAD2C*. a**, Oleic acid phenotypes for the years 2021 and 2022 compared to high-throughput marker genotyping for *AhFAD2A* and *AhFAD2B* loci. The double mutant *AhFAD2A* and *AhFAD2B* calls for high-oleic (>75%) lines are not consistent in the peanut MAGIC Core lines. Additionally, mutant *AhFAD2A* and wildtype *AhFAD2B* do not result in consistent oleic content between the founders TR (low-oleic) and WS16 (mid-oleic). The *AhFAD2C* C:ATTA haplotype was indicated as Mutant whereas C:A, T:ATTA, T:A were marked as wildtype 1-3, respectively. **b**, Pangenomic visualization of *AhFAD2C* highlighting promoter region variation and haplotypes across the founder lines. **c**, *AhFAD2C* MAGIC Core haplotypes compared to average oleic acid concentration. Individual phenotypes are represented as orange points. The *AhFAD2C* haplotype that all high-oleic founders possess is represented by a green bar. Error bars represent the standard error of the mean. Statistical significance was calculated using a Student's *t*-test. '**' represents a *P* value of $1.3x10^{-03}$ and '****' represents a *P* value of $4.6x10^{-05}$. 'ns' represents no statistical significance. Gray dashed lines represent 55% and 75% thresholds for mid and high-oleic phenotypes, respectively. **d**, I-VI, Alphafold2 predictions of a continuous open reading frame of *AhFAD2A* (I) and 180° rotation (II), *AhFAD2B* (III) and 180° rotation (VI), and *AhFAD2C* (V) and 180° rotation (VI) showing similarities in secondary amino acid structures. Green, orange, and blue regions highlight residues 50-100, 101-150, and 151-200, respectively.

Supplementary Table 3.1: Chromosome sizes for the assemblies of PeanutMAGIC founders 'Georgia-13M' (13M), 'SunOleic-97R' (97R), 'GT-C20' (C20), and 'Florida-07' (F07) with a comparison to the community reference 'Tifrunner' (TR).

Genotype	Chromosome	Total Size (Mb)	Total Size Compared to TR (Mb)
13M	13M.Chr.01	114.78	2.36
	13M.Chr.02	103.12	-0.18
	13M.Chr.03	144.27	1.16
	13M.Chr.04	123.86	-4.94
	13M.Chr.05	115.60	-0.94
	13M.Chr.06	121.98	3.00
	13M.Chr.07	81.50	-0.25
	13M.Chr.08	51.31	-0.22
	13M.Chr.09	122.16	1.66
	13M.Chr.10	119.72	2.64
	13M.Chr.11	151.05	1.76
	13M.Chr.12	119.62	-0.91
	13M.Chr.13	146.46	0.16
	13M.Chr.14	147.46	4.22
	13M.Chr.15	149.99	-10.04
	13M.Chr.16	150.13	-1.11
	13M.Chr.17	133.92	-0.27
	13M.Chr.18	136.82	1.79
	13M.Chr.19	160.20	0.84
	13M.Chr.20	146.22	1.19
97R	97R.Chr.01	114.98	2.56
	97R.Chr.02	100.99	-2.31
	97R.Chr.03	144.77	1.66
	97R.Chr.04	123.87	-4.93
	97R.Chr.05	110.43	-6.11
	97R.Chr.06	120.00	1.02
	97R.Chr.07	81.43	-0.32
	97R.Chr.08	52.25	0.72
	97R.Chr.09	121.67	1.17
	97R.Chr.10	119.36	2.28
	97R.Chr.11	149.05	-0.24
	97R.Chr.12	121.77	1.24
	97R.Chr.13	146.77	0.47

	97R.Chr.14	147.75	4.51
	97R.Chr.15	159.56	-0.47
	97R.Chr.16	149.51	-1.73
	97R.Chr.17	133.95	-0.24
	97R.Chr.18	136.88	1.85
	97R.Chr.19	160.29	0.93
	97R.Chr.20	148.02	2.99
C20	C20.Chr.01	114.81	2.39
	C20.Chr.02	101.52	-1.78
	C20.Chr.03	144.45	1.34
	C20.Chr.04	124.21	-4.59
	C20.Chr.05	108.99	-7.55
	C20.Chr.06	115.16	-3.82
	C20.Chr.07	81.10	-0.65
	C20.Chr.08	50.46	-1.07
	C20.Chr.09	119.64	-0.86
	C20.Chr.10	118.54	1.46
	C20.Chr.11	148.29	-1.00
	C20.Chr.12	119.71	-0.82
	C20.Chr.13	147.12	0.82
	C20.Chr.14	139.06	-4.18
	C20.Chr.15	159.33	-0.70
	C20.Chr.16	145.49	-5.75
	C20.Chr.17	133.32	-0.87
	C20.Chr.18	132.65	-2.38
	C20.Chr.19	159.60	0.24
	C20.Chr.20	144.49	-0.54
F07	F07.Chr.01	110.90	-1.52
	F07.Chr.02	104.28	0.98
	F07.Chr.03	141.17	-1.94
	F07.Chr.04	123.72	-5.08
	F07.Chr.05	107.19	-9.35
	F07.Chr.06	113.33	-5.65
	F07.Chr.07	79.95	-1.80
	F07.Chr.08	50.71	-0.82
	F07.Chr.09	125.27	4.77
	F07.Chr.10	112.16	-4.92
	F07.Chr.11	146.28	-3.01
	F07.Chr.12	120.17	-0.36
	F07.Chr.13	144.30	-2.00
	F07.Chr.14	140.66	-2.58

	F07.Chr.15	153.07	-6.96
	F07.Chr.16	147.78	-3.46
	F07.Chr.17	133.85	-0.34
	F07.Chr.18	132.60	-2.43
	F07.Chr.19	159.21	-0.15
	F07.Chr.20	144.19	-0.84

Supplementary Table 3.2: Chromosome sizes for the assemblies of PeanutMAGIC founders ‘NC94022’ (NC), ‘TifNV-High O/L’ (TNV), ‘Tifrunner’ (TR), and ‘GP-NC WS16’ (WS16) with a comparison to the community reference ‘Tifrunner’ (TR).

Genotype	Chromosome	Total Size (Mb)	Total Size Compared to TR (Mb)
NC	NC.Chr.01	116.00	3.58
	NC.Chr.02	106.49	3.19
	NC.Chr.03	146.10	2.99
	NC.Chr.04	129.91	1.11
	NC.Chr.05	119.62	3.08
	NC.Chr.06	130.35	11.37
	NC.Chr.07	88.74	6.99
	NC.Chr.08	65.38	13.85
	NC.Chr.09	125.94	5.44
	NC.Chr.10	121.39	4.31
	NC.Chr.11	149.64	0.35
	NC.Chr.12	122.50	1.97
	NC.Chr.13	146.91	0.61
	NC.Chr.14	145.44	2.20
	NC.Chr.15	153.76	-6.27
	NC.Chr.16	150.23	-1.01
	NC.Chr.17	135.64	1.45
	NC.Chr.18	140.17	5.14
	NC.Chr.19	159.92	0.56
	NC.Chr.20	147.69	2.66
TNV	TNV.Chr.01	115.79	3.37
	TNV.Chr.02	104.18	0.88
	TNV.Chr.03	146.83	3.72
	TNV.Chr.04	136.81	8.01
	TNV.Chr.05	113.42	-3.12
	TNV.Chr.06	126.03	7.05
	TNV.Chr.07	89.46	7.71
	TNV.Chr.08	60.93	9.40
	TNV.Chr.09	150.55	30.05
	TNV.Chr.10	126.08	9.00
	TNV.Chr.11	149.76	0.47
	TNV.Chr.12	122.36	1.83
	TNV.Chr.13	146.70	0.40

	TNV.Chr.14	145.89	2.65
	TNV.Chr.15	159.92	-0.11
	TNV.Chr.16	150.75	-0.49
	TNV.Chr.17	139.47	5.28
	TNV.Chr.18	138.30	3.27
	TNV.Chr.19	159.73	0.37
	TNV.Chr.20	146.61	1.58
TR	TR.Chr.01	112.42	-
	TR.Chr.02	103.30	-
	TR.Chr.03	143.11	-
	TR.Chr.04	128.80	-
	TR.Chr.05	116.54	-
	TR.Chr.06	118.98	-
	TR.Chr.07	81.75	-
	TR.Chr.08	51.53	-
	TR.Chr.09	120.50	-
	TR.Chr.10	117.08	-
	TR.Chr.11	149.29	-
	TR.Chr.12	120.53	-
	TR.Chr.13	146.30	-
	TR.Chr.14	143.24	-
	TR.Chr.15	160.03	-
	TR.Chr.16	151.24	-
	TR.Chr.17	134.19	-
	TR.Chr.18	135.03	-
	TR.Chr.19	159.36	-
	TR.Chr.20	145.03	-
WS16	WS16.Chr.01	114.59	2.17
	WS16.Chr.02	107.14	3.84
	WS16.Chr.03	137.01	-6.10
	WS16.Chr.04	134.27	5.47
	WS16.Chr.05	118.57	2.03
	WS16.Chr.06	127.98	9.00
	WS16.Chr.07	91.89	10.14
	WS16.Chr.08	42.19	-9.34
	WS16.Chr.09	125.70	5.20
	WS16.Chr.10	123.32	6.24
	WS16.Chr.11	136.33	-12.96
	WS16.Chr.12	115.85	-4.68
	WS16.Chr.13	146.52	0.22
	WS16.Chr.14	147.73	4.49

	WS16.Chr.15	153.30	-6.73
	WS16.Chr.16	150.67	-0.57
	WS16.Chr.17	145.96	11.77
	WS16.Chr.18	136.92	1.89
	WS16.Chr.19	159.91	0.55
	WS16.Chr.20	147.95	2.92

Supplementary Table 3.3: PeanutMAGIC founders' assembly data.

Genome	Number of Contigs	Total Assembly Length (Mb)	Contig N50 (Mb)
13M	102	2540	55
97R	94	2543	57
C20	111	2507	52
F07	238	2490	21
NC	180	2601	27
TNV	90	2629	60
TR	3496	2538	1
WS16	103	2563	49

Supplementary Table 3.4: PeanutMAGIC founders' BUSCO scores.

Genome	BUSCO (%)	Single Copy BUSCO (%)	Duplicated BUSCO (%)	Fragmented (%)	Missing (%)
13M	97	7.7	89.3	0.2	2.8
97R	96.7	7	89.7	0.2	3.1
C20	96.6	7.5	89.1	0.3	3.1
F07	95.2	5.9	89.3	0.3	4.5
NC	97	6.8	90.2	0.2	2.8
TNV	97	7.1	89.9	0.2	2.8
TR	97.6	4.3	93.3	0.2	2.2
WS16	97.6	16.1	81.5	0.2	2.2

Supplementary Table 3.5: Comparison of PeanutMAGIC Pangenome homeologous chromosome number of variants.

	all	SNP	Non-SNP
Chr.01vChr.11	153392	30288	123104
Chr.02vChr.12	47015	35675	11340
Chr.03vChr.13	106878	47142	59736
Chr.04vChr.14	989	628	361
Chr.05vChr.15	2882	5325	2443
Chr.06vChr.16	1898	4906	6804
Chr.07vChr.17	1029	1924	895
Chr.08vChr.18	37728	28018	9710

Supplementary Table 3.6: Annotated genes within the significant region on Chr.05 from the single reference marker oleic acid average GWAS. The region identified spanned from 90,216,715 to 90,219,762 (3 Kb) on Chr.05.

Chromosome	Start	End	Gene	Annotation
TR.Chr.05	90,002,709	90,004,511	arahy.RX4XEC	Pentatricopeptide repeat (PPR) superf
TR.Chr.05	90,006,239	90,009,505	arahy.IMTK78	Pentatricopeptide repeat (PPR) superf
TR.Chr.05	90,009,612	90,010,244	arahy.K7FXBU	Pentatricopeptide repeat (PPR) superf
TR.Chr.05	90,016,993	90,018,568	arahy.5QN4SS	Pentatricopeptide repeat (PPR) superf
TR.Chr.05	90,025,677	90,027,440	arahy.FH5YY3	Pentatricopeptide repeat (PPR) superf
TR.Chr.05	90,037,372	90,037,626	arahy.6450JK	Unknown protein
TR.Chr.05	90,038,639	90,041,182	arahy.YDZR5J	protein n=1 Tax=Oryza sativa subs RepID=C7J641_ORYSJ
TR.Chr.05	90,041,272	90,043,034	arahy.62B9ZV	CTV.20 n=1 Tax=Poncirus tri RepID=Q8H6Q8_PONTI
TR.Chr.05	90,043,126	90,043,961	arahy.LJ6V76	viral movement protein
TR.Chr.05	90,055,167	90,059,112	arahy.RKDW8Y	ATP binding/protein serine/threonine k max]
TR.Chr.05	90,059,759	90,062,611	arahy.W1MG13	Putative methyltransferase famil
TR.Chr.05	90,063,441	90,071,153	arahy.P0W1U8	alcohol dehydrogenase 1
TR.Chr.05	90,067,310	90,071,125	arahy.586MWM	UPF0415 protein C7orf25 homolog [Glycine max]
TR.Chr.05	90,078,896	90,079,099	arahy.EHW2CP	2-hydroxyacyl-CoA lyase-like [Gl
TR.Chr.05	90,079,543	90,080,795	arahy.KBQ0HF	2-hydroxyacyl-CoA lyase-like [Gl
TR.Chr.05	90,103,228	90,106,224	arahy.2J0XRF	7SK snRNA methylphosphate cappin Tax=Boreoeutheria RepID=G3I65
TR.Chr.05	90,120,967	90,128,058	arahy.1EPI1N	Disease resistance protein (TIR-NBS-LI
TR.Chr.05	90,137,426	90,144,450	arahy.HVB0T8	disease resistance protein (TIR-NBS putative
TR.Chr.05	90,145,037	90,149,934	arahy.W1U3RA	Endosomal targeting BRO1-like doma protein
TR.Chr.05	90,157,084	90,162,288	arahy.JNVS5M	DNA replication factor CDT1-lik
TR.Chr.05	90,165,601	90,177,031	arahy.68FRR6	Disease resistance protein (TIR-NBS-LI
TR.Chr.05	90,177,791	90,183,747	arahy.MS9PXH	Disease resistance protein (TIR-NBS-LI
TR.Chr.05	90,183,922	90,190,624	arahy.WVJ4GZ	Disease resistance protein (TIR-NBS-LI
TR.Chr.05	90,206,002	90,208,634	arahy.PJ1Z6Q	disease resistance protein (TIR-NBS putative
TR.Chr.05	90,231,580	90,234,945	arahy.BXSS78	disease resistance protein (TIR-NBS putative
TR.Chr.05	90,235,420	90,241,327	arahy.WKI3HS	disease resistance protein (TIR-NBS putative
TR.Chr.05	90,251,350	90,251,682	arahy.GLN17G	Disease resistance protein (TIR-NBS-LI
TR.Chr.05	90,251,716	90,252,096	arahy.ZH17KP	TMV resistance protein N-like [Gl
TR.Chr.05	90,253,327	90,254,302	arahy.UT82H5	Disease resistance protein (TIR-NBS-LI

TR.Chr.05	90,288,686	90,292,152	arahy.M38AUV	disease resistance protein (TIR-NBS putative
TR.Chr.05	90,293,574	90,297,983	arahy.BXGY4D	TMV resistance protein N-like [Gl
TR.Chr.05	90,298,580	90,303,704	arahy.5MQ2SU	Disease resistance protein (TIR-NBS-LI
TR.Chr.05	90,309,451	90,311,467	arahy.9KKX2T	dentin sialophosphoprotein-like isofor max]
TR.Chr.05	90,314,218	90,317,586	arahy.KQ9JXS	TMV resistance protein N-like [Gl
TR.Chr.05	90,323,638	90,335,696	arahy.F56ZG1	Disease resistance protein (TIR-NBS-LI
TR.Chr.05	90,354,270	90,358,951	arahy.82EEBY	disease resistance protein (TIR-NBS putative
TR.Chr.05	90,361,118	90,366,748	arahy.JGCP84	beta glucosidase 13
TR.Chr.05	90,381,350	90,392,960	arahy.4Z6VDH	beta glucosidase 15
TR.Chr.05	90,450,724	90,451,813	arahy.3ZGA1D	putative ribonuclease H protein At1 [Glycine max]
TR.Chr.05	90,453,052	90,456,674	arahy.X8X2MM	uncharacterized protein LOC1004999 [Glycine max]

Supplementary Table 3.7: PeanutMAGIC founders' linear genome and *AhFAD2C* sequence alignment. '*' indicates a SNP location where the first base is the reference and the second is the query. '–' indicates a deletion in the query. The calls made by the aligner have been extracted into independent columns for readability.

Genome	Location	*AhFAD2C* Minimap call	SNP call	Indel call
13M	13M.Chr.19:153821845	cs:Z::2992*ct:167-tta:2063	C	Present
97R	97R.Chr.19:153899361	cs:Z::2992*ct:167-tta:2063	C	Present
C20	C20.Chr.19:153846211	cs:Z::2992*ct:2230	C	Absent
F07	F07.Chr.19:153442397	cs:Z::2992*ct:167-tta:2063	C	Present
NC	NC.Chr.19:154095551	cs:Z::3160-tta:2063	T	Present
TNV	TNV.Chr.19:153905630	cs:Z::2992*ct:167-tta:2063	C	Present
TR	TR.Chr.19:153531624	cs:Z::5223	T	Absent
WS16	WS16.Chr.19:154061793	cs:Z::2992*ct:167-tta:2063	C	Present

Supplementary Table 3.8: ANOVA statistics of *AhFAD2A*, *AhFAD2B,* and *AhFAD2C* genotypes as a function of oleic acid content in the PeanutMAGIC Core. Significant codes: '***' is <0.001 and ' ' is >1.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
AhFAD2A	1	7516	7516	99.005	<2E-16	***
AhFAD2B	1	15754	15754	207.518	<2E-16	***
AhFAD2C	3	2761	920	12.121	1.91E-07	***
AhFAD2A*:*AhFAD2C	2	53	26	0.347	0.707483	
AhFAD2B*:*AhFAD2C	2	7	3	0.046	0.955405	
AhFAD2A*:*AhFAD2B	1	926	926	12.197	0.000563	***
AhFAD2A*:*AhFAD2B*:*AhFAD2C	1	91	91	1.2	0.274306	
Residuals	257	19510	76			

Supplementary Table 3.9: ANOVA statistics of *AhFAD2A*, *AhFAD2B*, and *AhFAD2C* genotypes as a function of oleic acid content in the 'S' population. Significant codes: '***' is <0.001, '**' is <0.01, '.' is <0.1, and ' ' is >1.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
AhFAD2A	1	2555	2555	31.268	1.64E-06	***
AhFAD2B	1	3694	3694	45.169	4.06E-08	***
AhFAD2C	1	1004	1004	12.285	1.12E-03	**
AhFAD2A*:*AhFAD2C	1	286	286	3.499	0.06856	.
AhFAD2A*:*AhFAD2B	1	229	229	2.808	0.10141	
Residuals	41	3351	82			

Supplementary Table 3.10: Tukey's HSD statistics of *AhFAD2A*, *AhFAD2B*, and *AhFAD2C* genotypes as a function of oleic acid content in the PeanutMAGIC population.

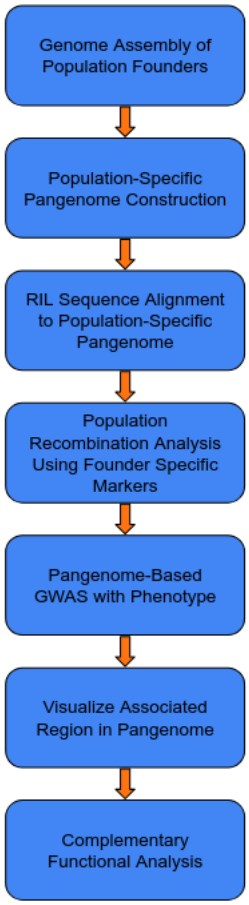
	Genotype	diff	lwr	upr	p adj
AhFAD2A	G-A	-12.47	-14.96	-9.98	0
AhFAD2B	CT-C	15.35	13.22	17.48	0
AhFAD2C	C:ATTA-C:A	7.42	3.64	11.19	$4.3x10^{-6}$
	T:A-C:A	-8.84	-18.11	0.44	0.068
	T:ATTA-C:A	-9.83	-25.75	6.09	0.38
	T:A-C:ATTA	-16.26	-26.04	-6.47	$1.4x10^{-4}$
	T:ATTA-C:ATTA	-17.25	-33.47	-1.02	0.032
	T:ATTA-T:A	-0.99	-19.29	17.31	1.00

Supplementary Table 3.11: Tukey's HSD statistics of *AhFAD2A*, *AhFAD2B*, and *AhFAD2C* genotypes as a function of oleic acid content in the 'S' population.

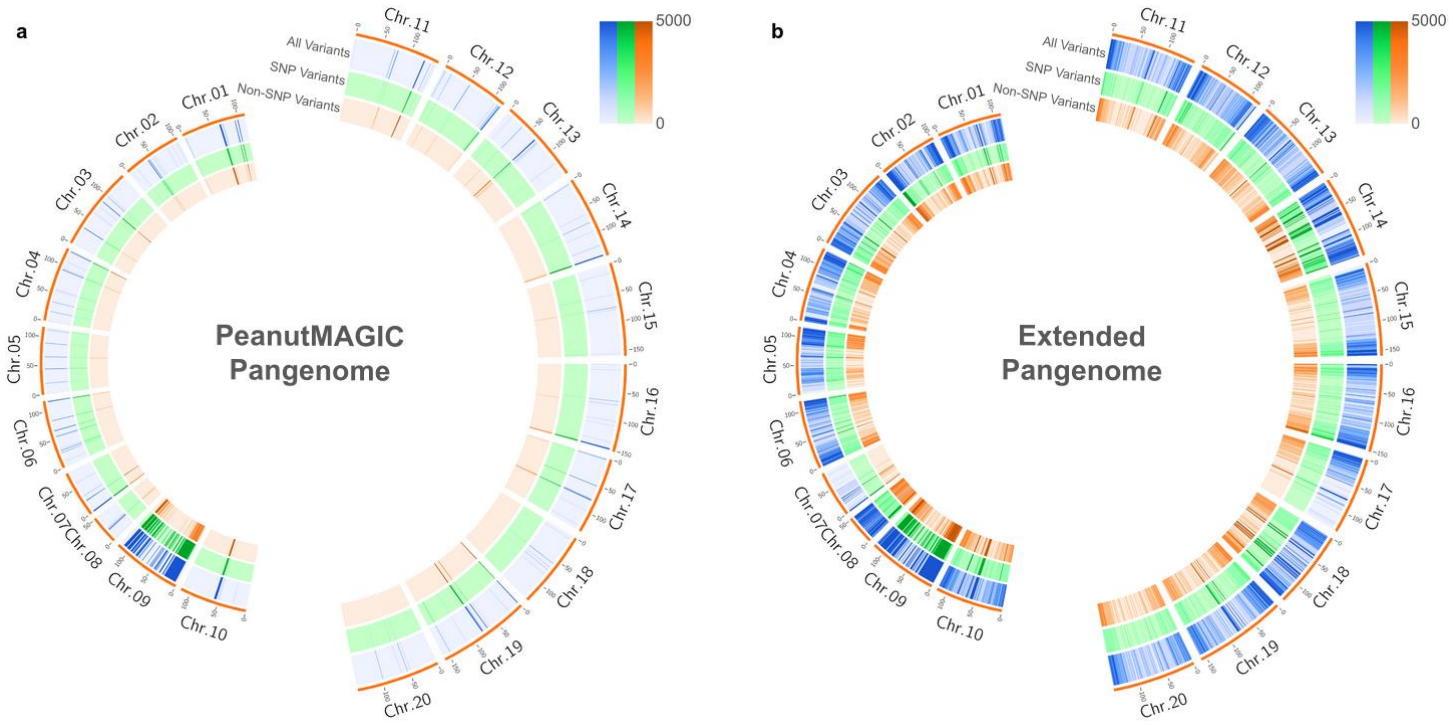
	Genotype	diff	lwr	upr	p adj
AhFAD2A	G-A	-14.51	-20.15	-8.87	$5.5x10^{-6}$
AhFAD2B	CT-C	17.15	11.54	22.76	$2.0x10^{-7}$
AhFAD2C	T-C	-16.56	-26.56	-6.57	$1.7x10^{-3}$

Supplementary Table 3.12: Short read low coverage data for *AhFAD2C* genotypes compared to long read *AhFAD2C* genotypes of the six mutant *AhFAD2A* and *AhFAD2B* lines with mid-oleic phenotypes.

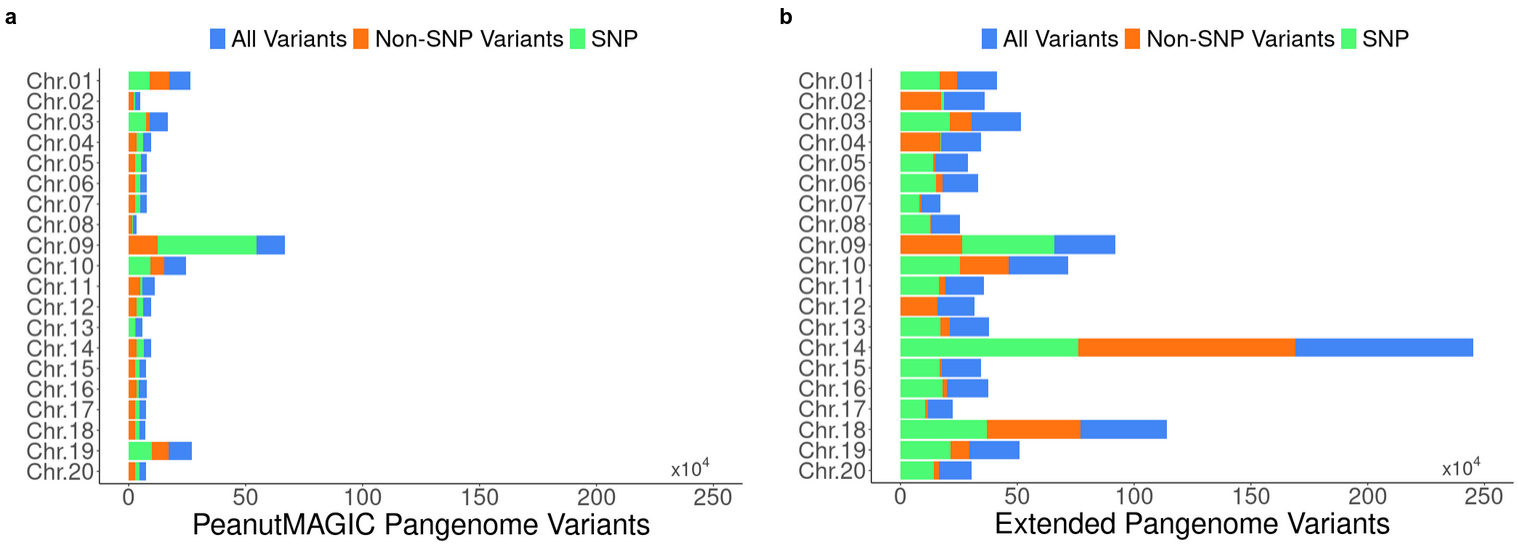
Genotype	Short Read *AhFAD2C*	Long Read *AhFAD2C*
MG1201	-	wildtype1
MG14	wildtype1	-
MG1404	wildtype2	mutant
MG2004	wildtype3	wildtype3
MG2405	-	-
MG906	wildtype3	wildtype3



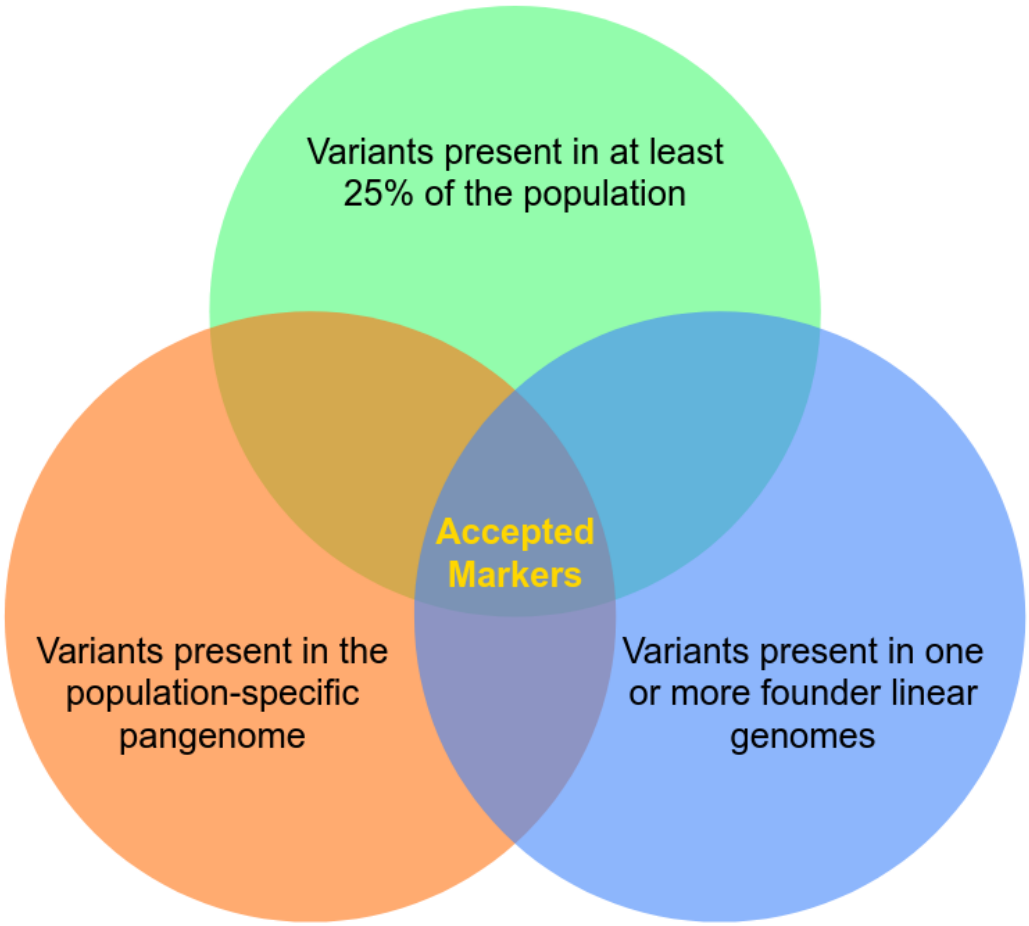
Supplementary Fig. 3.1: Visualization of Study Workflow. This study starts from the assembly of the founder genomes of PeanutMAGIC to facilitate the construction of the PeanutMAGIC pangenome. The population specific pangenome offers a comprehensive library of variants in the population. RIL sequences were aligned to the pangenome to generate pangenome based markers that have founder origin information. The markers were used to analyze recombination patterns within the population for genomic characterization to enable association studies. The markers were then used to associate oleic acid. The associated regions were examined and visualized using the population-specific pangenome to identify variants of interest and compared to recombination patterns. The isolated variants were examined in complementary studies to validate the findings of the associations.



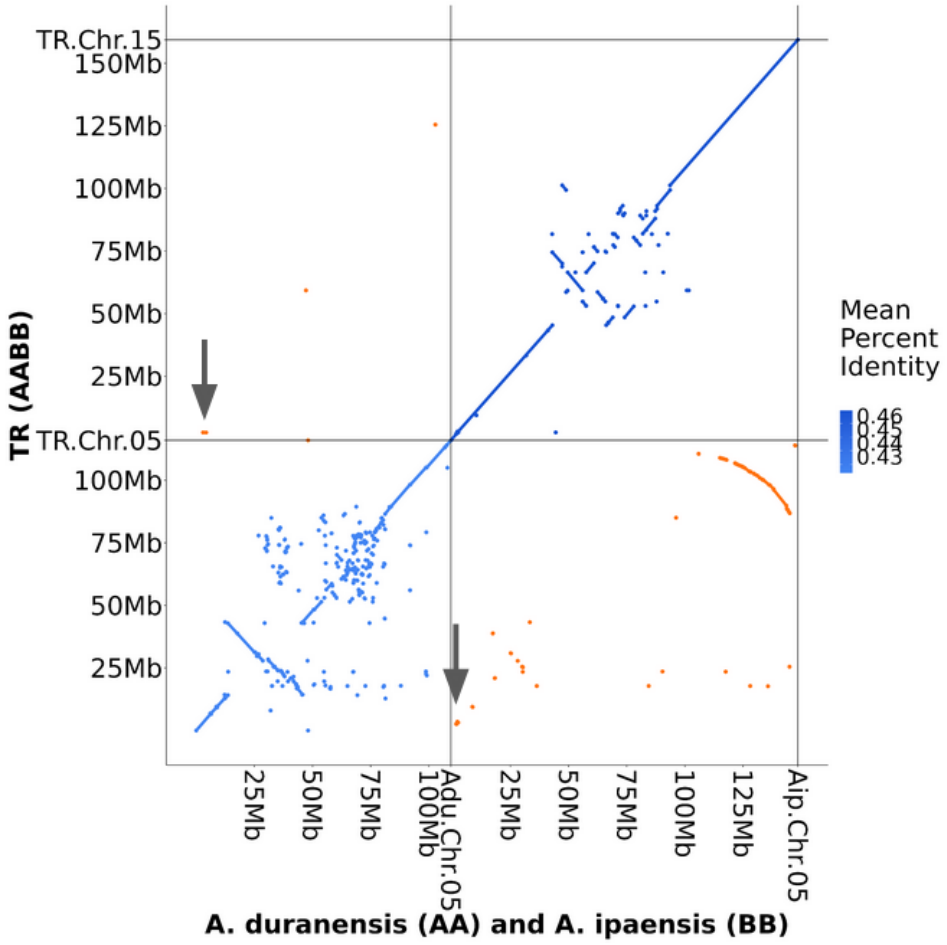
Supplementary Fig. 3.2: a, Circos plot representing A (Chr.01-Chr.10, left side) and B (Chr.11-Chr.20, right side) subgenomes of cultivated peanut with outer orange bars as scales in Mb. The distribution of all variation sites within the PeanutMAGIC pangenome is represented as a blue heatmap (per 1 Mb), SNP variant locations are represented as a green heatmap (per 1 Mb), and non-SNP variants are represented as an orange heatmap (per 1 Mb). **b**, Circos plot representing A (Chr.01-Chr.10, left side) and B (Chr.11-Chr.20, right side) subgenomes of cultivated peanut with outer orange bars as scales in Mb. The distribution of all variation sites within the extended pangenome is represented as a blue heatmap (per 1 Mb), SNP variant locations are represented as a green heatmap (per 1 Mb), and non-SNP variants are represented as an orange heatmap (per 1 Mb).



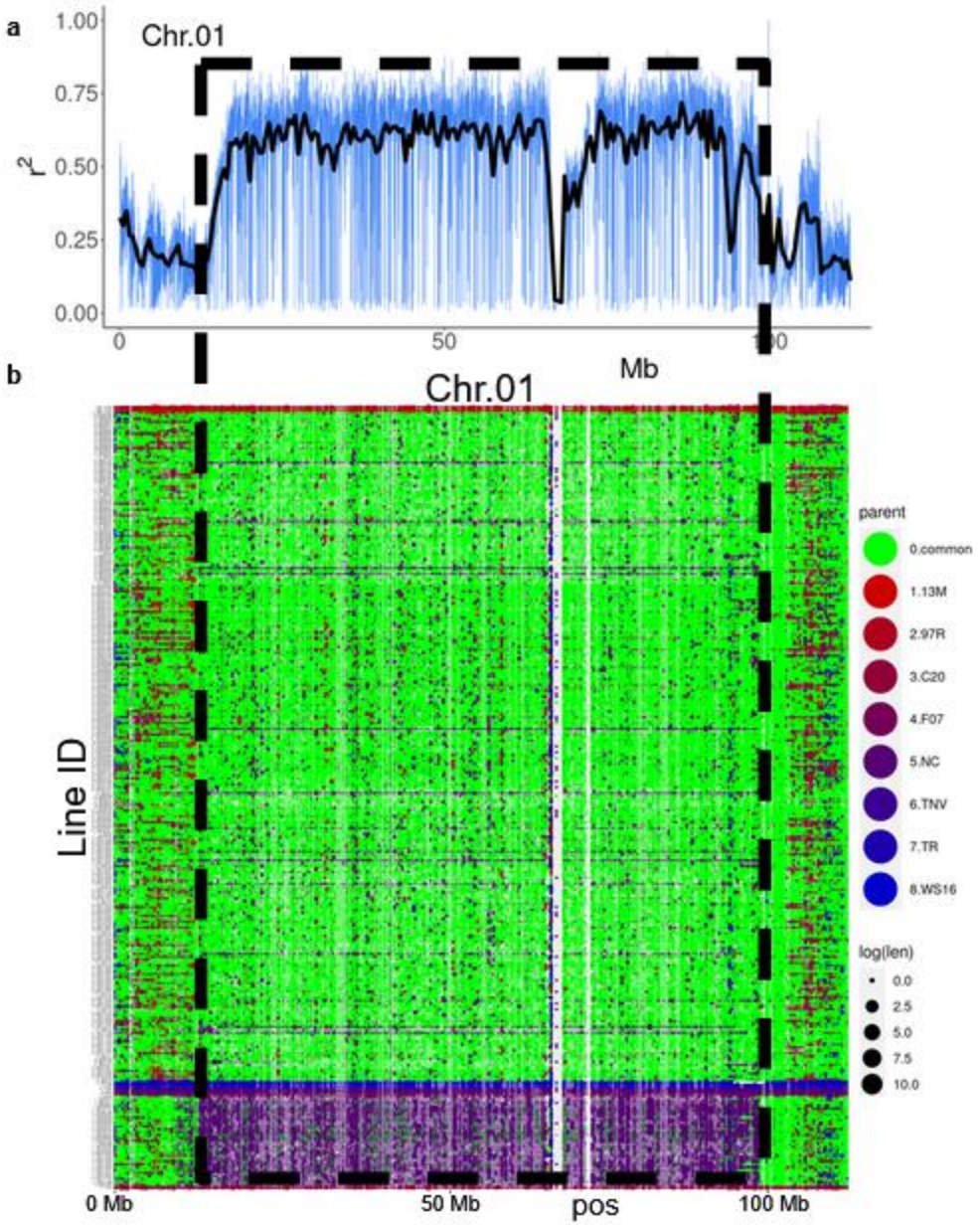
Supplementary Fig. 3.3: a, Number of PeanutMAGIC pangenome variants per chromosome. Blue bars represent all variants, orange bars represent non-SNP variants, and green bars represent SNP variants. Note: parts of the longer bars are overshadowed by smaller bars. **b**, Number of the extended pangenome variants per chromosome. Blue bars represent all variants, orange bars represent non-SNP variants, and green bars represent SNP variants.



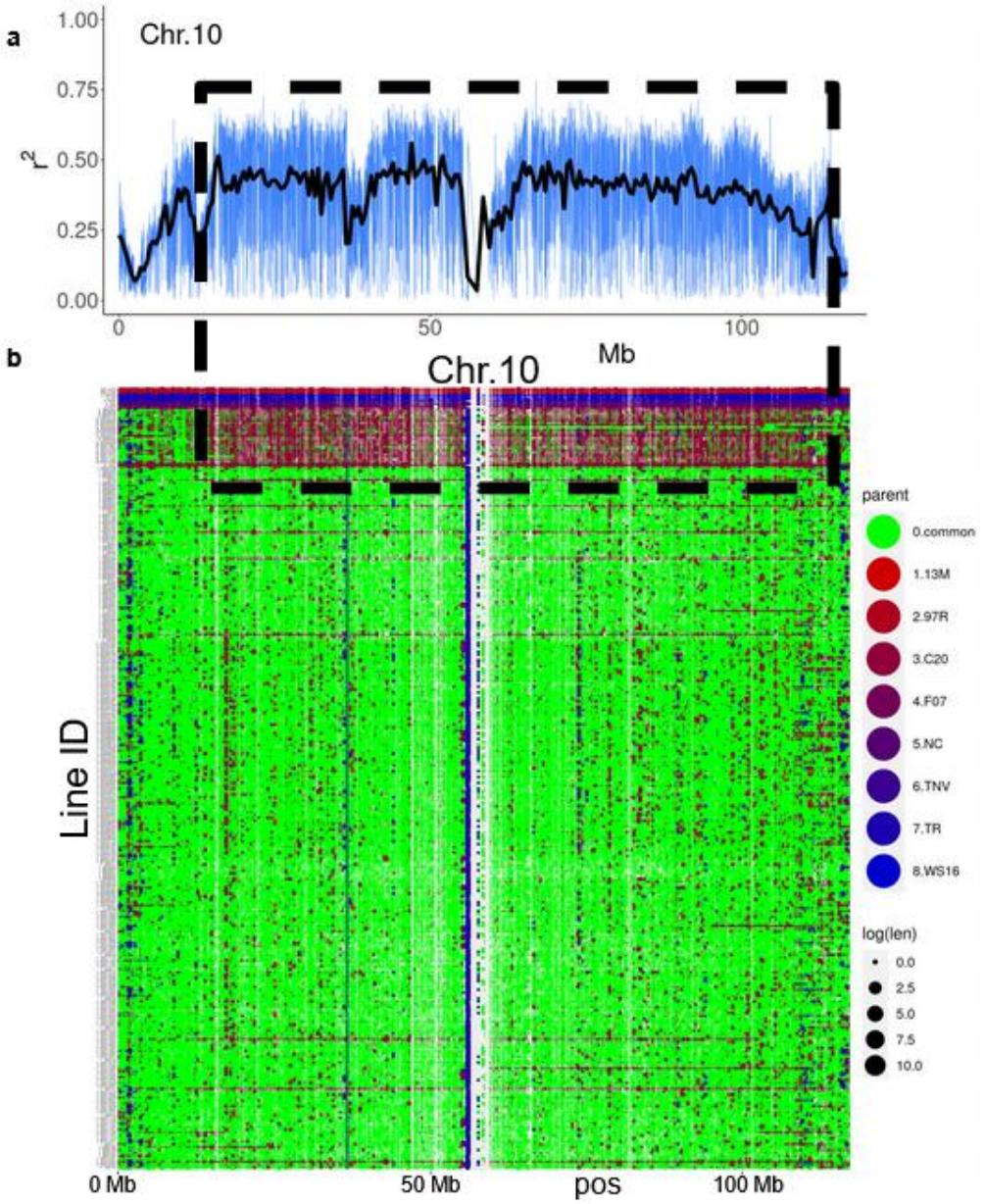
Supplementary Fig. 3.4: Venn diagram representing the three criteria that a variant must possess for it to be considered a marker for analysis of the MAGIC Core and the 'S' populations.



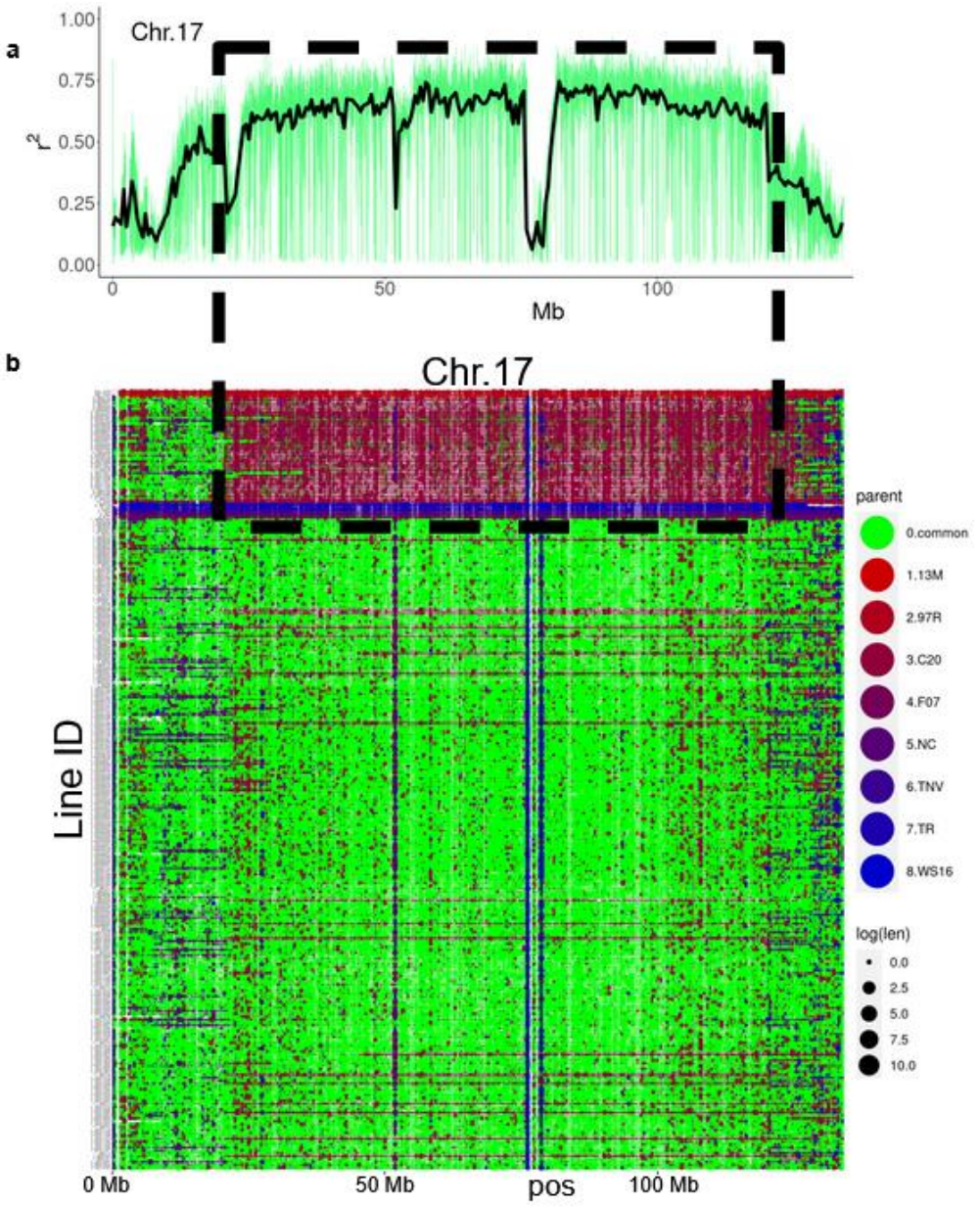
Supplementary Fig. 3.5: Alignment of Chr.05 and Chr.15 of TR (AABB) to *A. duranensis* (Adu, AA) and *A. ipaensis* (Aip, BB). Blue represents on-target mapping (AA-AA or BB-BB) and orange represents off-target (AA-BB or BB-AA) mapping. Gray arrows point to a homeologous region that contains AA and BB signatures, indicating this region retained fragments from either subgenome during homeologous exchange.



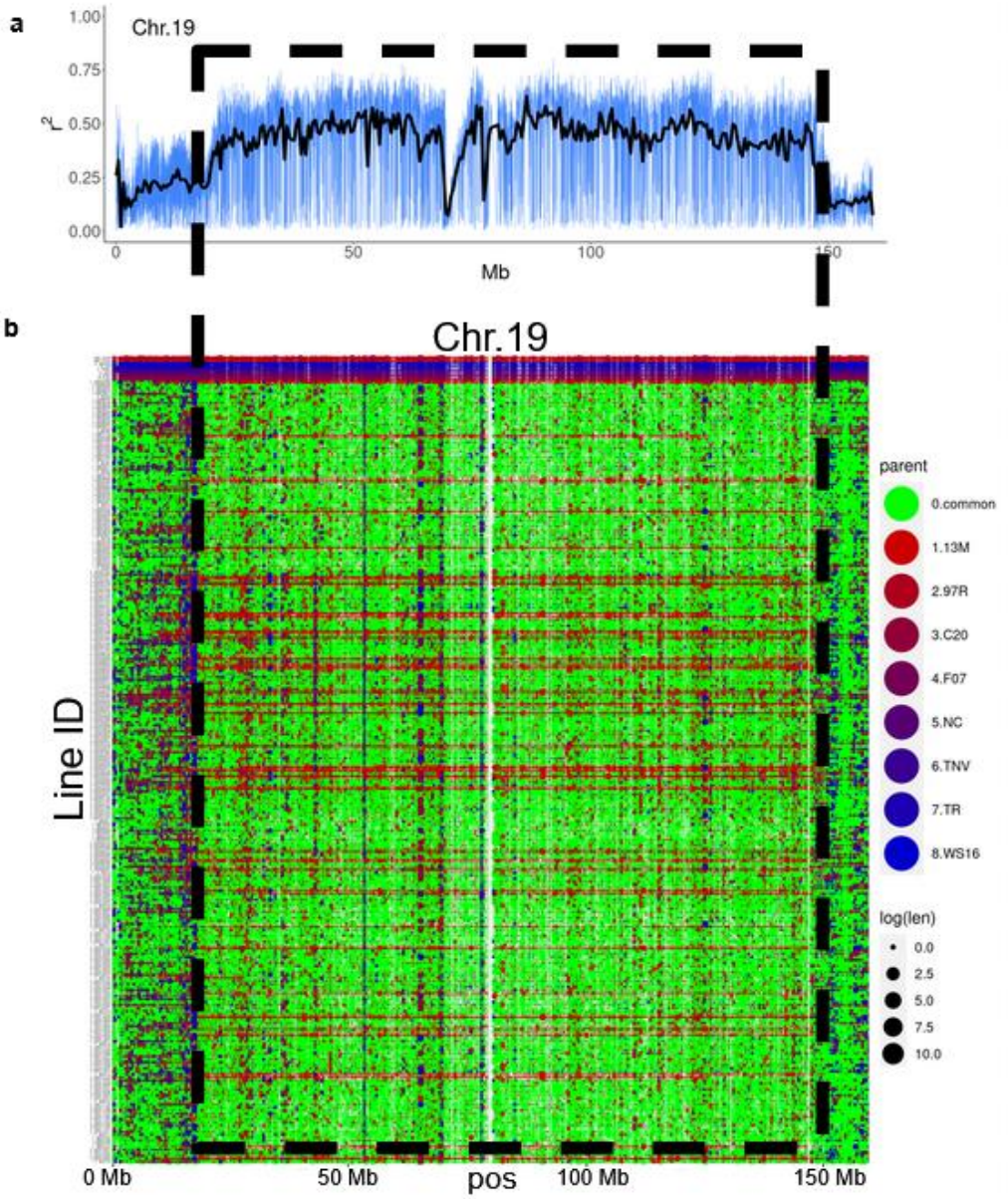
Supplementary Fig. 3.6: Characteristics of Chr.01 recombination. **a**, r^2 values across Chr.01 calculated from a sliding window of 300 markers, depicting the pericentromeric region with elevated LD. The black line represents the average r^2 value in 0.5 Mb. **b**, PeanutMAGIC Core pangenome marker calls for Chr.01 with founder origin coloration. The black dashed box highlights the pericentromeric region with increased LD stemming from NC.



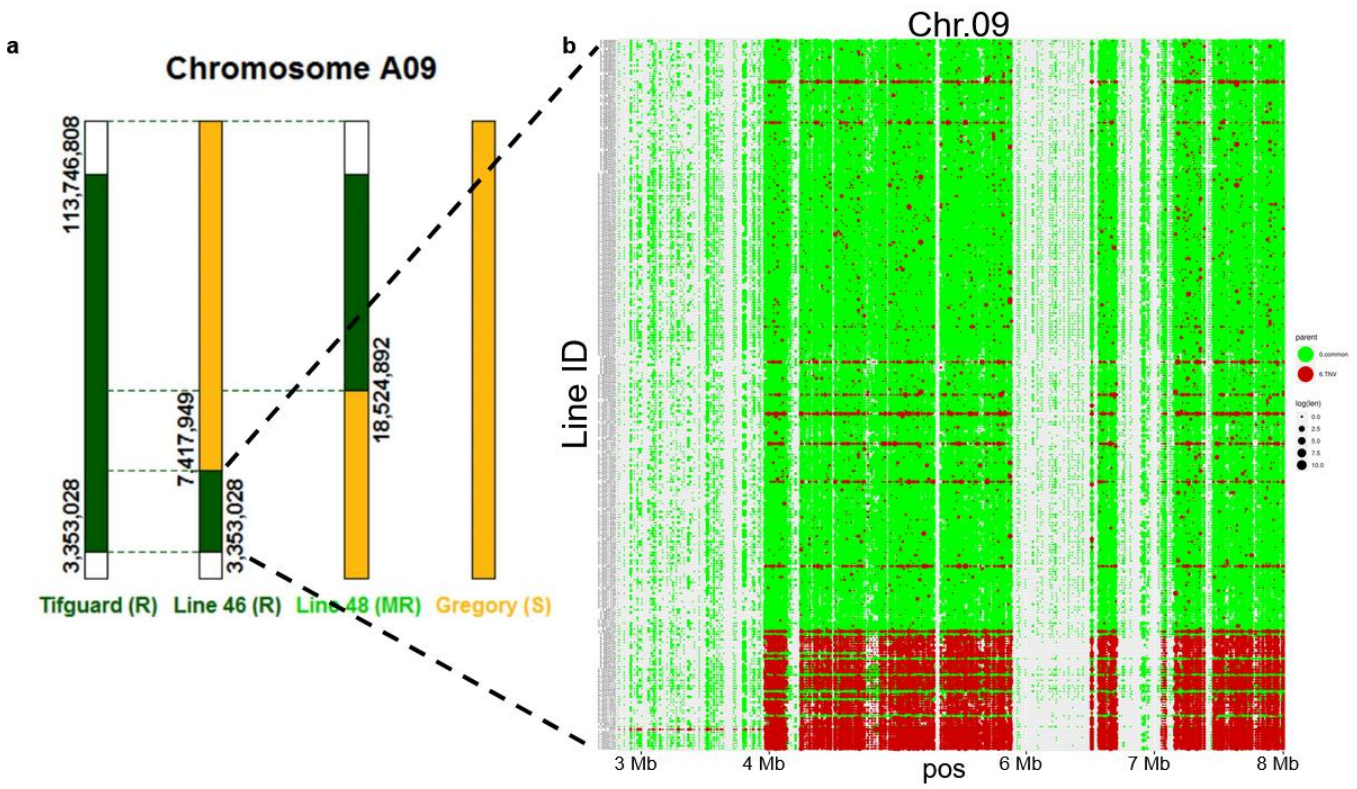
Supplementary Fig. 3.7: Characteristics of Chr.10 recombination. **a**, r^2 values across Chr.10 calculated from a sliding window of 300 markers depicting the pericentromeric region with elevated LD. The black line represents the average r^2 value in 0.5 Mb. **b**, PeanutMAGIC Core pangenome marker calls for Chr.10 with founder origin coloration. The black dashed box highlights the pericentromeric region with increased LD stemming from C20.



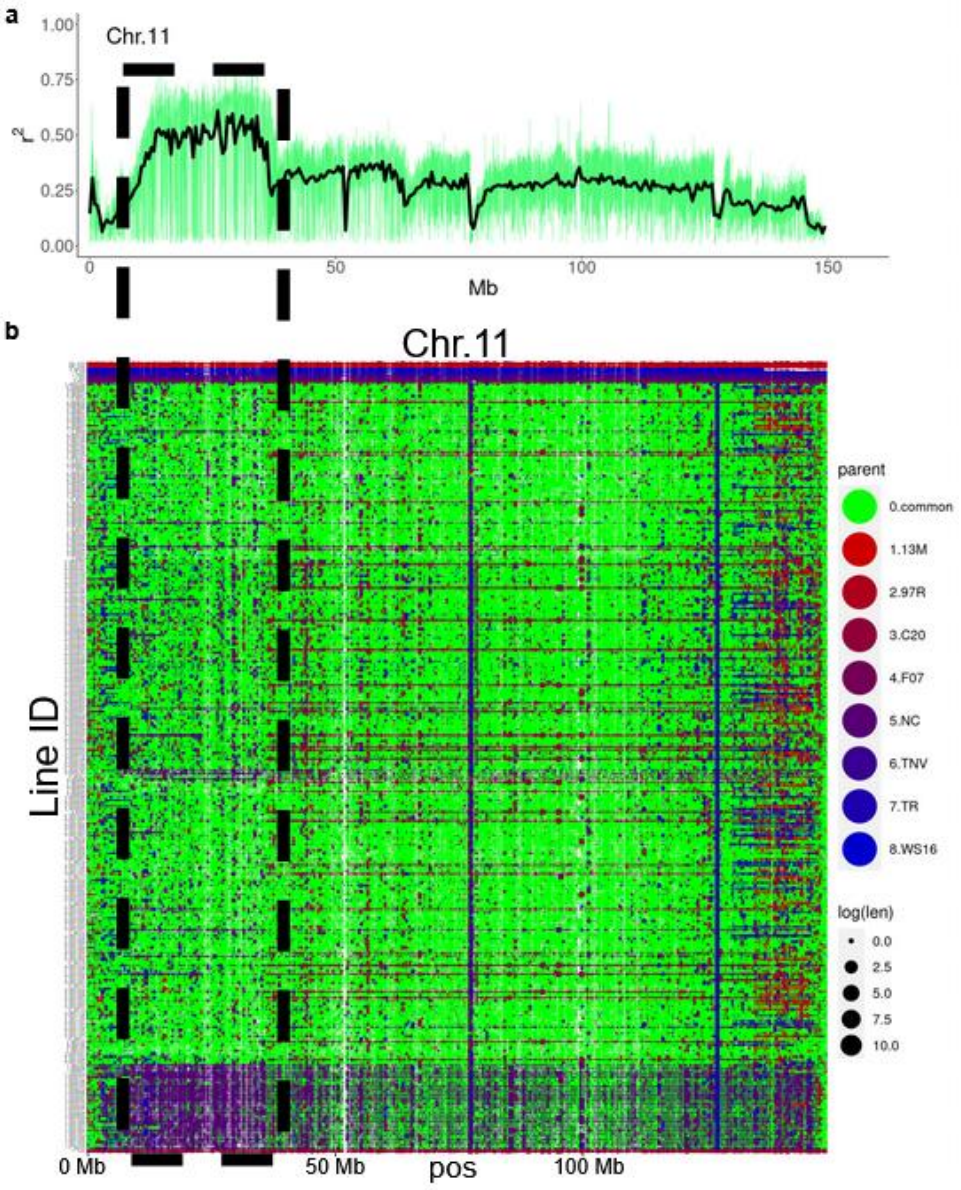
Supplementary Fig. 3.8: Characteristics of Chr.17 recombination. **a**, r^2 values across Chr.17 calculated from a sliding window of 300 markers depicting the pericentromeric region with elevated LD. The black line represents the average r^2 value in 0.5 Mb. **b**, PeanutMAGIC Core pangenome marker calls for Chr.17 with founder origin coloration. The black dashed box highlights the pericentromeric region with increased LD stemming from C20.



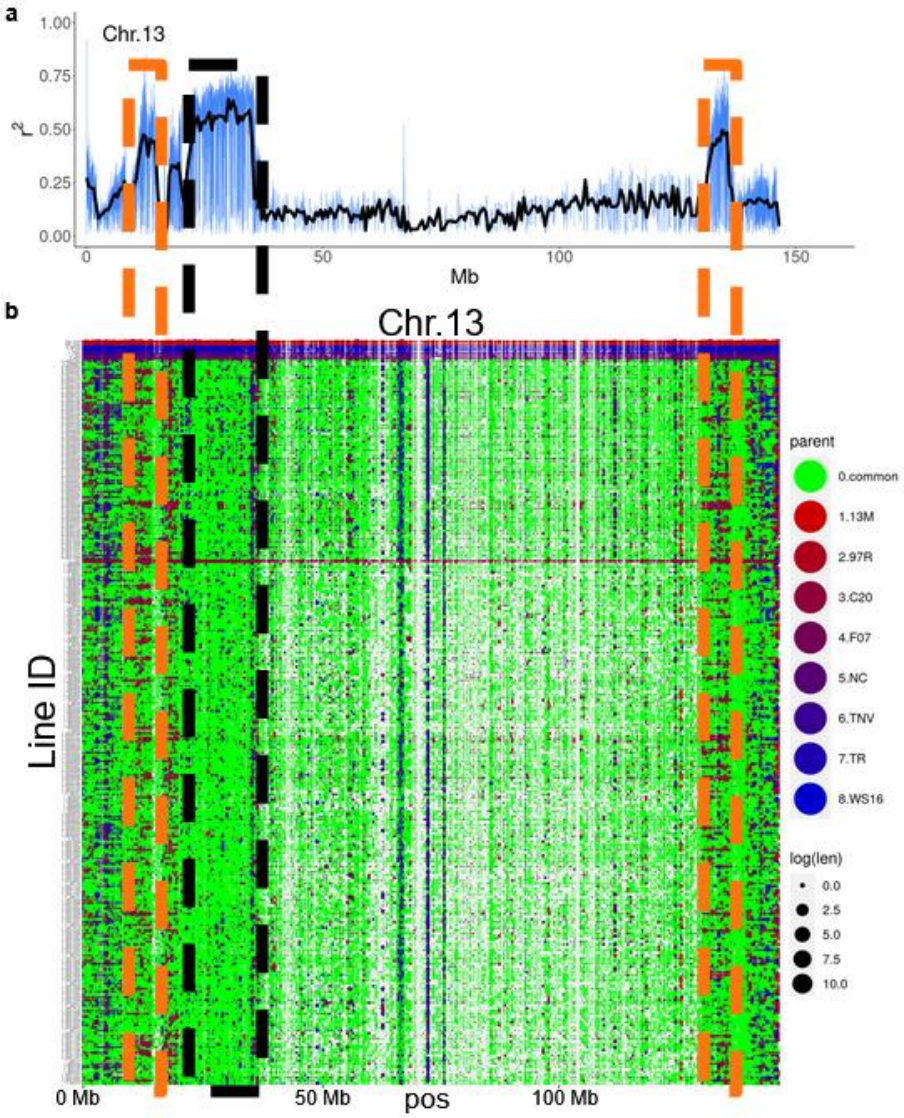
Supplementary Fig. 3.9: Characteristics of Chr.19 recombination. **a**, r^2 values across Chr.19 calculated from a sliding window of 300 markers depicting the pericentromeric region with elevated LD. The black line represents the average r^2 value in 0.5 Mb. **b**, PeanutMAGIC Core pangenome marker calls for Chr.19 with founder origin coloration. The black dashed box highlights the pericentromeric region with increased LD stemming from 13M.



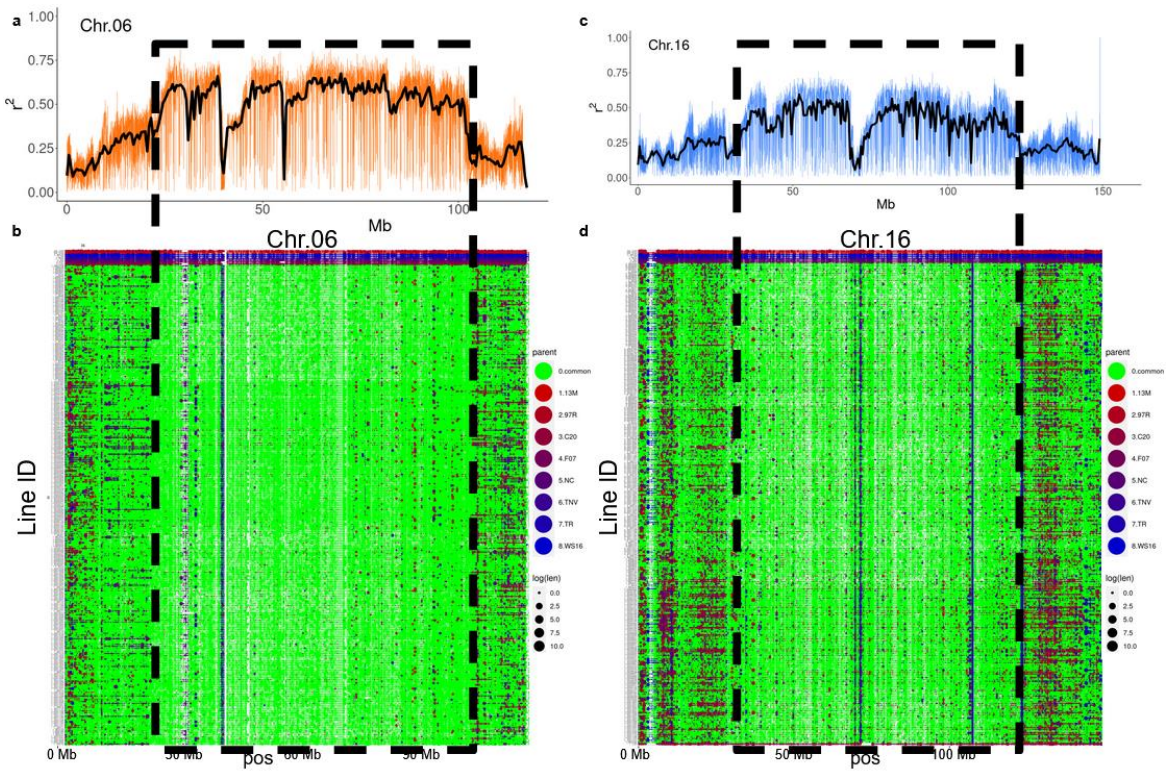
Supplementary Fig. 3.10: Comparison of *A. cardenasii* recombination in a biparental population and PeanutMAGIC Core. **a**, Figure 1 from Clevenger et al. (2017), depicting the region of *A. cardenasii* introgression that confers resistance to peanut root-knot nematode. **b**, A zoom-in of the region identified by Clevenger et al. (2017), using the PeanutMAGIC Core and pangenome markers, showing an increase in marker density and micro-recombinants from TNV, the donor of *A. cardenasii* alleles in PeanutMAGIC. This is visualized by the intermixed green markers within solid red lines, depicting locations that were once from the introgression but have been replaced by genomic information from another founder and *vice versa*. Gray regions indicate areas without marker data.



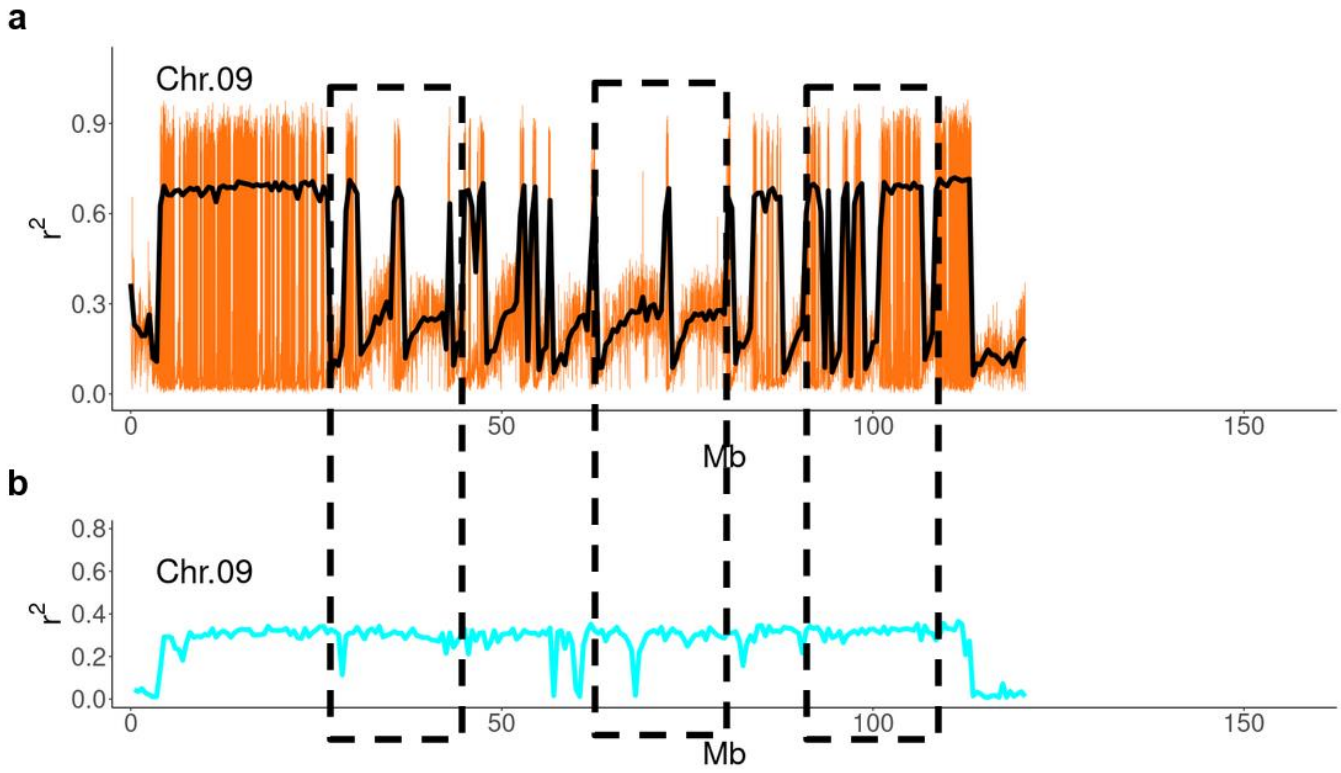
Supplementary Fig. 3.11: Characteristics of Chr.11 recombination. **a**, r^2 values across Chr.11 calculated from a sliding window of 300 markers depicting a region with elevated LD. The black line represents the average r^2 value in 0.5 Mb. **b**, PeanutMAGIC Core pangenome marker calls for Chr.11 with founder origin coloration. The black dashed box highlights a region with increased LD stemming from NC in the beginning of Chr.11. This region recombined less than the middle and end regions of Chr.11 from NC.



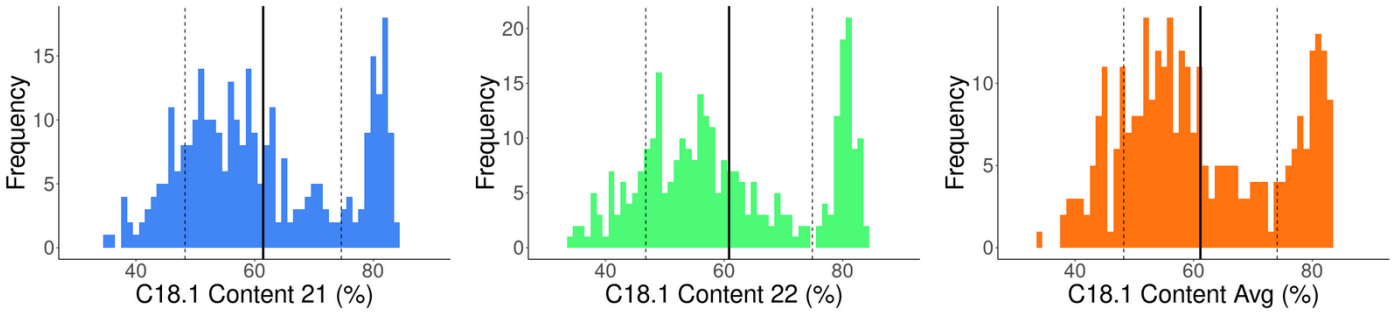
Supplementary Fig. 3.12: Characteristics of Chr.13 recombination. **a**, r^2 values across Chr.13 calculated from a sliding window of 300 markers depicting three regions with elevated LD. The black line represents the average r^2 value in 0.5 Mb. **b**, PeanutMAGIC Core pangenome marker calls for Chr.13 with founder origin coloration. The orange dashed boxes highlight two small regions from C20 that segregated as small whole blocks, increasing LD in these regions. The black dashed box highlights a region with low marker diversity, hindering the ability to track recombination and increasing LD.



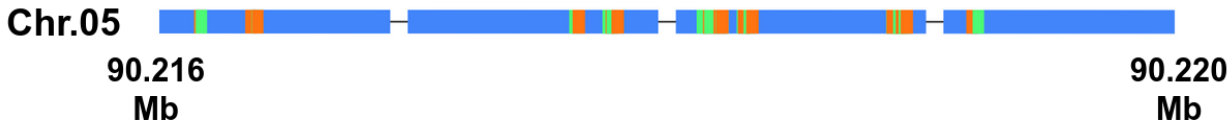
Supplementary Fig. 3.13: Characteristics of homeologous Chr.06 and Chr.16 recombination. **a**, r^2 values across Chr.06 calculated from a sliding window of 300 markers depicting the pericentromeric region with elevated LD. The black line represents the average r^2 value in 0.5 Mb. **b**, PeanutMAGIC Core pangenome marker calls for Chr.06 with founder origin coloration. The black dashed box highlights a region with low marker diversity, reducing the ability to track recombination within the region. **c**, r^2 values across Chr.16 calculated from a sliding window of 300 markers depicting the pericentromeric region with elevated LD. The black line represents the average r^2 value in 0.5 Mb. **d**, PeanutMAGIC Core pangenome marker calls for Chr.16 with founder origin coloration. The black dashed box highlights a region with low marker diversity, reducing the ability to track recombination within the region. The difficulty in track recombination of these regions increases the region's LD due to the lack of detecting recombination and does not indicate that recombination did not occur in these regions.



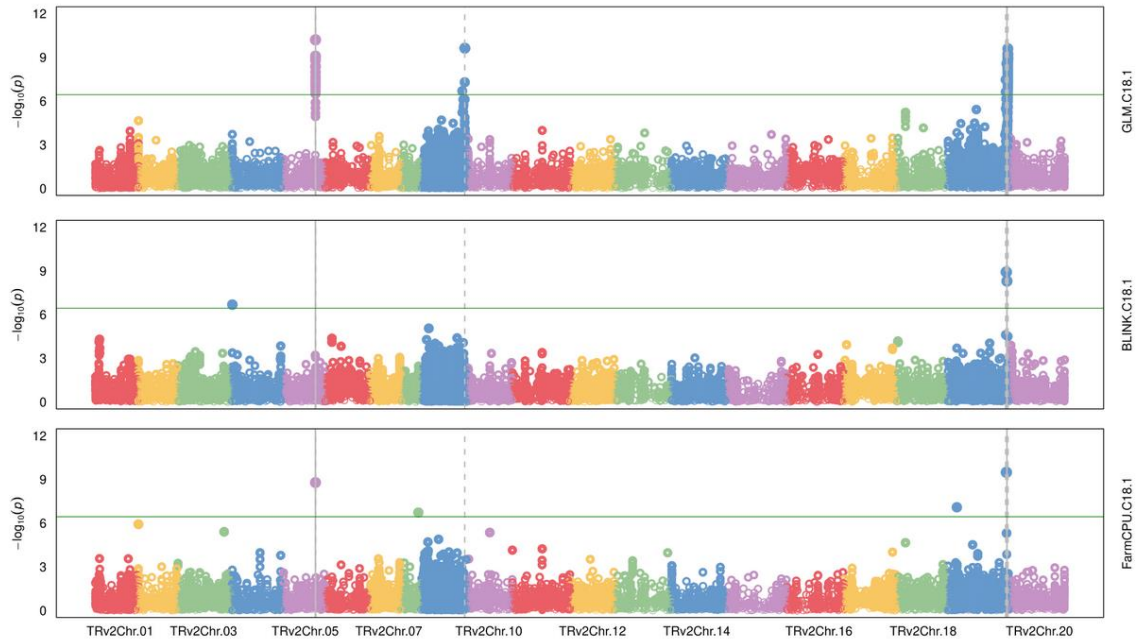
Supplementary Fig. 3.14: Comparison of Chr.09 recombination with single reference and PeanutMAGIC based markers. **a**, r^2 values across Chr.09 calculated from a sliding window of 300 pangenome-based markers. The black line represents the average r^2 value in 0.5 Mb. **b**, Figure 4b from Thompson et al. (2024) depicting the r^2 values across Chr.09 calculated from a sliding window of 100 single reference markers and averaged in 0.5 Mb bins. The increased number of markers in the sliding window for PeanutMAGIC pangenome markers is due to the increased marker density, allowing for equal physical window sizes between single reference and pangenome-based markers. The black dashed box highlights regions that were previously determined to have consistent LD using the single reference markers but have variable LD using the PeanutMAGIC pangenome based markers, offering an improved understanding of the recombination in this region in the PeanutMAGIC Core.



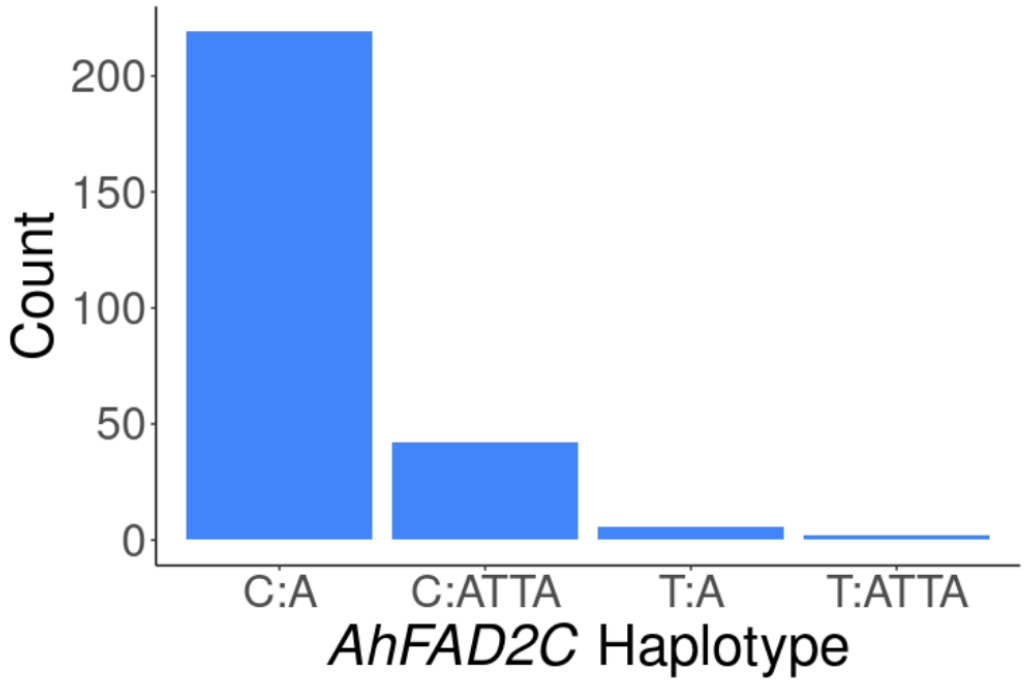
Supplementary Fig. 3.15: Oleic acid content for the 2021 season (blue), the 2022 season (green), and the average for an individual between 2021 and 2022 (orange). The black solid vertical line represents the PeanutMAGIC Core average for the respective phenotype, and the dashed vertical line represents the average plus or minus one standard deviation, respectively.



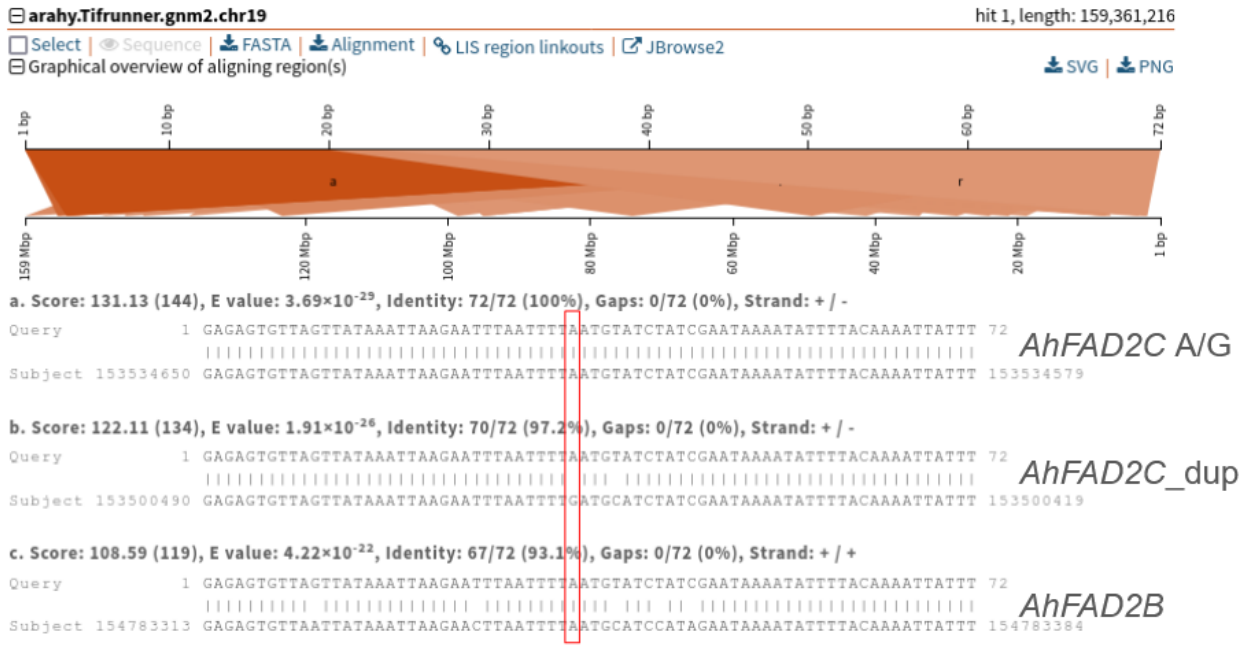
Supplementary Fig. 3.16: PeanutMAGIC pangenome visualization of the single-reference oleic acid associated region on Chr.05, spanning 3 Kb (90,216,715- 90,219,762). There are 30 single-reference-based markers within the region and are highlighted by orange and green bars that are highlighted in 50 bp blocks. Note some blocks are overlapped with other blocks. These regions do not align with variants in the pangenome and there are no variants in the pangenome for the associated region.



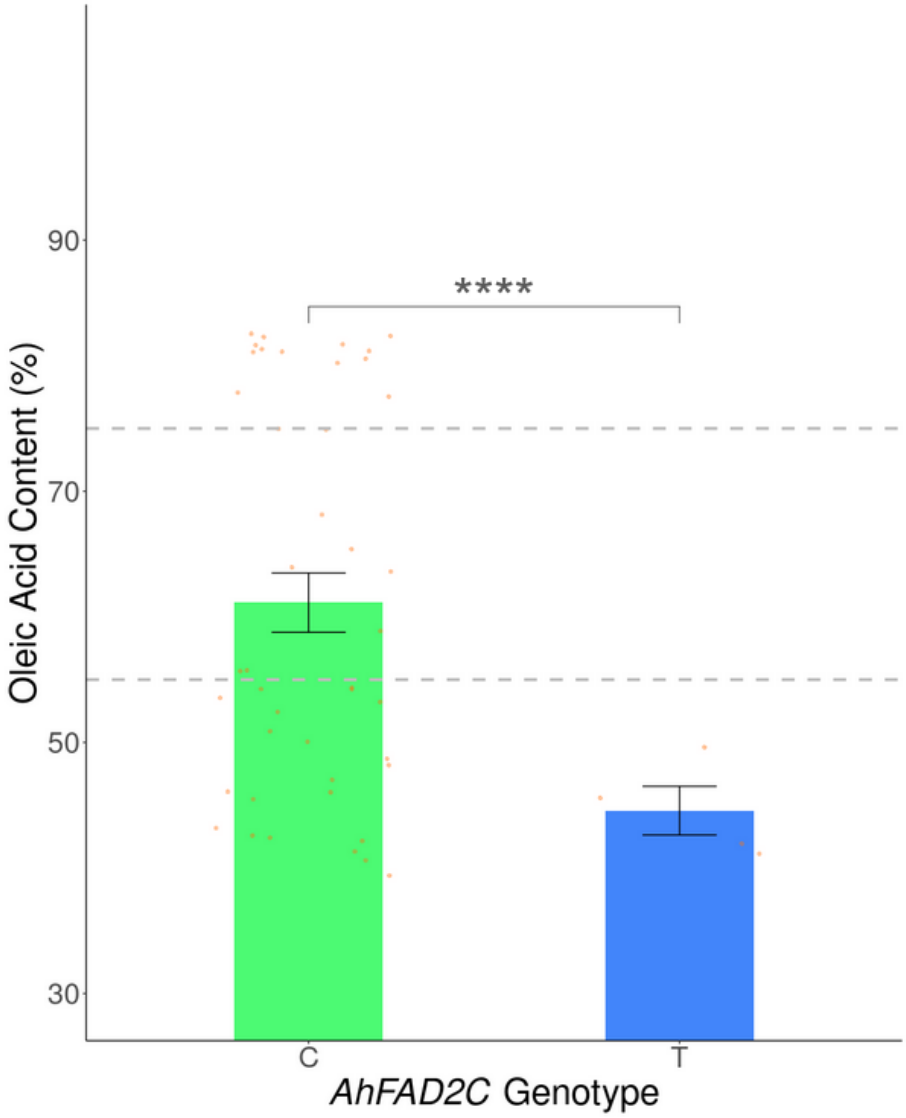
Supplementary Fig. 3.17: Single-reference-based marker associations using generalized linear model (GLM), fixed and random model circulating probability unification (FarmCPU), and Bayesian-information and linkage-disequilibrium iteratively nested keyway (BLINK) models. The models FarmCPU and BLINK consistently identify Chr.19, however, miss the known functional gene *AhFAD2A* on Chr.09 and identify regions that are not consistent with previous studies.



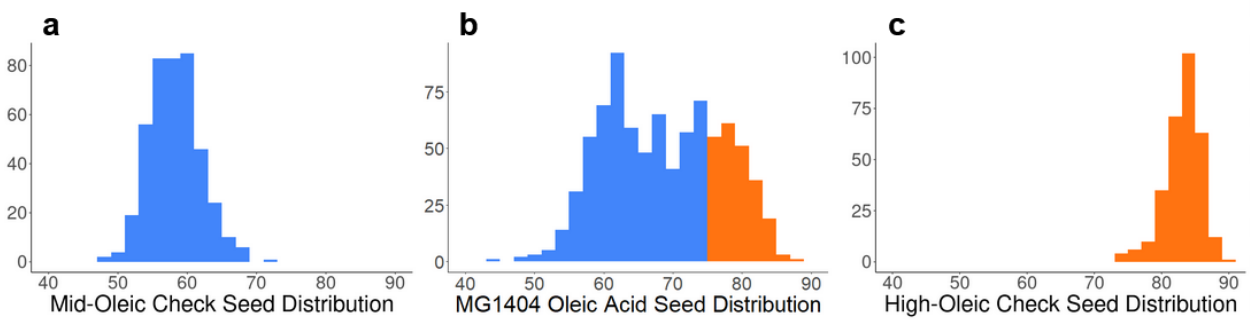
Supplementary Fig. 3.18: Distribution of *AhFAD2C* haplotypes in the MAGIC Core population.



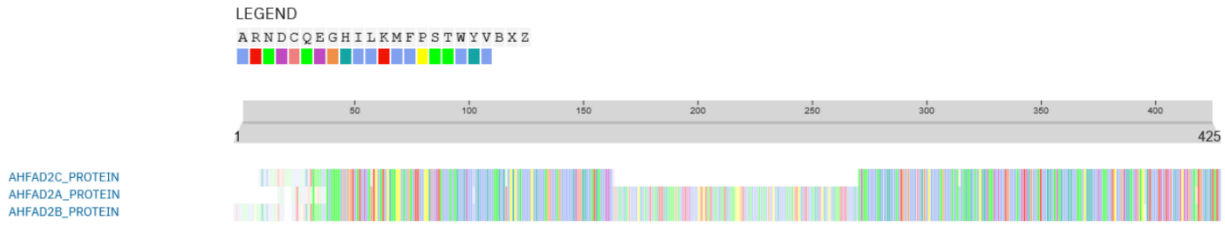
Supplementary Fig. 3.19: Alignment of *AhFAD2C* SNP sequence to TR showing multiple sites with similar sequence throughout the chromosome in orange. Of the similar sequence regions included a duplicated promoter region of *AhFAD2C* (*AhFAD2C*_dup) and *AhFAD2B*. The red box highlights the location of the C/T SNP (+ strand) or A/G SNP (- strand).



Supplementary Fig 3.20: *AhFAD2C* 'S' population genotypes compared to the average oleic acid concentration. Individual phenotypes are represented as orange points. The *AhFAD2C* genotype of the high oleic founder 'SunOleic 97R' (97R) is highlighted green. Error bars represent the standard error of the mean. Statistical significance was calculated using a Student's *t*-test. **** represents a *P* value of $5.7x10^{-05}$. Gray dashed lines represent 55% and 75% thresholds for mid and high oleic phenotypes, respectively.



Supplementary Fig 3.21: a, Oleic acid distribution for the mid-oleic acid check. **b**, Oleic acid distribution for the line MG1404 displaying both mid-oleic and high-oleic seed. **c**, Oleic acid distribution for high-oleic acid check.



Supplementary Fig 3.22: Amino acid sequence alignment for AhFAD2A, AhFAD2B, and AhFAD2C. Generally, the sequence is conserved among the three genes, except for a large deletion in *AhFAD2C* compared to *AhFAD2A* and *AhFAD2B*, which shortens the length of the *AhFAD2C* amino acid sequence.

Supplementary Appendix

KhufuPAN bootstrapping

⇒ Deconstruct GFA > VCF

⇒ Filter for high QUAL

⇒ Filter out monomorphic variants

⇒ Filter out variants that are only polymorphic with the reference

⇒ Filter out variants having odd nucleotides

⇒ Filter out multi-call variants (only keep variants that are mapped to a single location)

⇒ Optional: filter out variants with misaligning of the borders

⇒ Getting the final parental calls (Filtered-Variant set)

KhufuVAR processing

⇒ FASTQ files run through Khufu-core standards (https://www.hudsonalpha.org/khufudata/)

⇒ Modified FASTQ files are mapped by Giraffe (https://github.com/vgteam/vg/wiki/Mapping-short-reads-with-Giraffe)

⇒ Variants are called batches and are aligned with the Filtered-Variant set

⇒ Missing data across samples, and the Filtered-Variant set, and MAF are calculated.

⇒ Combining the batches in a single dataset

⇒ Filter out variants >= MAF

⇒ Filter out variants <= missing %

⇒ Filter out samples <= missing %

⇒ Getting the final panmap/fasta pair

File formats

Filtered-Variant set

chr_pos	13M,97R,C20,F07,NC,TNV,TR,WS16	13M,97R,C20,F07,NC,TNV,TR,WS16	unique_alleles	len
TRv2Chr.01_2	AAA,,,,,AC,AC,AAC	1,0,0,0,0,3,3,2	AAA,AAC,AC	3,3,2
TRv2Chr.01_7	T,,,,,A,A,A	2,0,0,0,0,1,1,1	A,T	1,1
TRv2Chr.01_11	CG,,,,CCTA,CCTA,CCTA,CCTA	2,0,0,0,1,1,1,1	CCTA,CG	4,2
TRv2Chr.01_18	CGTT,,,,C,C,C,C	2,0,0,0,1,1,1,1	C,CGTT	1,4
TRv2Chr.01_20	CG,,,,TA,TA,TA,TA	1,0,0,0,2,2,2,2	CG,TA	2,2
TRv2Chr.01_26	CGTT,,,,CTAAA,CTAA,CTAAA,CTAAA	1,0,0,0,3,2,3,3	CGTT,CTAA,CTAAA	4,4,5
TRv2Chr.01_32	A,,,,C,C,C,C	1,0,0,0,2,2,2,2	A,C	1,1
TRv2Chr.01_34	A,,,,T,T,T,T	1,0,0,0,2,2,2,2	A,T	1,1
TRv2Chr.01_38	G,,,,C,C,A,C	3,0,0,0,2,2,1,2	A,C,G	1,1,1

Column 1: Chromosome Position

Column 2: Comma-delimited parental variants sequence

Column 3: Comma-delimited allele IDs of parental calls

Column 4: Comma-delimited list of unique alleles

Column 5: Comma-delimited list of length of unique alleles

*.panmap

chr	pos	len	13M,97R,C20,F07,NC,TNV,TR,WS16	MG2751	MG2508	MG2753	MG2607	MG2707	MG2703
TRv2Chr.01	76282	1,3,15	1,1,0,1,2,1,3,1	-	3	-	3	3	3
TRv2Chr.01	102614	1,1	2,2,1,2,2,2,2,2	1	-	2	2	2	2
TRv2Chr.01	147505	1,1	1,1,2,1,1,1,2,1	2	1	1	1	1	2
TRv2Chr.01	164480	1,1	1,1,2,1,1,1,1,1	2	1	1	-	1	1
TRv2Chr.01	167689	1,2	2,2,1,2,2,2,2,2	1	2	2	2	2	-
TRv2Chr.01	190978	1,1	2,2,1,2,2,2,2,2	1	2	2	2	2	2
TRv2Chr.01	191706	1,2	2,2,1,2,2,2,2,2	1	2	2	2	2	2
TRv2Chr.01	192108	1,1	2,2,1,2,2,2,2,2	1	2	2	2	2	2
TRv2Chr.01	192820	1,1	2,2,1,2,2,2,2,2	1	2	2	2	2	2

Column 1: Chromosome

Column 2: Position of variant start site

Column 3: Comma-delimited length of the alleles (alleles IDs are in a sequential order)

Column 4: Comma-delimited allele IDs of the parental calls

Column 5-n: Sample/Column calls using allele IDs

*.panmap.fasta

Allele sequences in fasta formats

Sequence Names are following chr_pos_id

```
>TRv2Chr.01_76282_1
G
>TRv2Chr.01_76282_2
GTC
>TRv2Chr.01_76282_3
GTCTCTCTCTCTC
>TRv2Chr.01_102614_1
A
>TRv2Chr.01_102614_2
G
```

*.panmap_heatmap.txt

Calls are referenced based on the parents

chr	pos	len	13M,97R,C20,F07,NC,TNV,TR,WS16	MG2751	MG2508	MG2753	MG2607	MG2707	MG2703
TRv2Chr.01	76282	1,3,15	1,1,0,1,2,1,3,1	0	7	0	7	7	7
TRv2Chr.01	102614	1,1	2,2,1,2,2,2,2,2	3	0	1,2,4,5,6,7,8	1,2,4,5,6,7,8	1,2,4,5,6,7,8	1,2,4,5,6,7,8
TRv2Chr.01	147505	1,1	1,1,2,1,1,1,2,1	3,7	1,2,4,5,6,8	1,2,4,5,6,8	1,2,4,5,6,8	1,2,4,5,6,8	3,7
TRv2Chr.01	164480	1,1	1,1,2,1,1,1,1,1	3	1,2,4,5,6,7,8	1,2,4,5,6,7,8	0	1,2,4,5,6,7,8	1,2,4,5,6,7,8
TRv2Chr.01	167689	1,2	2,2,1,2,2,2,2,2	3	1,2,4,5,6,7,8	1,2,4,5,6,7,8	1,2,4,5,6,7,8	1,2,4,5,6,7,8	0
TRv2Chr.01	190978	1,1	2,2,1,2,2,2,2,2	3	1,2,4,5,6,7,8	1,2,4,5,6,7,8	1,2,4,5,6,7,8	1,2,4,5,6,7,8	1,2,4,5,6,7,8
TRv2Chr.01	191706	1,2	2,2,1,2,2,2,2,2	3	1,2,4,5,6,7,8	1,2,4,5,6,7,8	1,2,4,5,6,7,8	1,2,4,5,6,7,8	1,2,4,5,6,7,8
TRv2Chr.01	192108	1,1	2,2,1,2,2,2,2,2	3	1,2,4,5,6,7,8	1,2,4,5,6,7,8	1,2,4,5,6,7,8	1,2,4,5,6,7,8	1,2,4,5,6,7,8
TRv2Chr.01	192820	1,1	2,2,1,2,2,2,2,2	3	1,2,4,5,6,7,8	1,2,4,5,6,7,8	1,2,4,5,6,7,8	1,2,4,5,6,7,8	1,2,4,5,6,7,8

*.panmap_heatmap.rds

The binary format of *.panmap_heatmap.txt can be shown on Rshiny interface (https://w-korani.shinyapps.io/cyclops_eye_ii/)

CHAPTER 4

POPULATION-SPECIFIC PANGENOMES ENABLE THE DISCOVERY OF A COPY NUMBER VARIANT FOR TSWV RESISTANCE IN PEANUT

[1] Thompson, E. Population-Specific Pangenomes Enable the Discovery of a Copy Number Variant for TSWV Resistance in Peanut. To be Submitted to *Nature Genetics*.

Abstract

Cultivated peanut (*Arachis hypogea* L.) is an important legume crop that provides a rich source of protein and oil for humans globally. *Tomato Spotted Wilt Virus* (TSWV) severely limits peanut production and is managed through resistant cultivars. Here we utilized population-specific pangenome-based genotypes to identify TSWV resistance from 'NC94022' in PeanutMAGIC and the 'S' biparental population. Population specific pangenomics revealed a copy number variant (CNV) from 'NC94022' within the associated region of both populations, containing tandem duplications of a novel glutamate receptor gene. Using short read sequence data, we created presence absence variation calls for the CNV and found it significantly influenced TSWV severity across both populations. However, we explored the difficulties of short read sequences to retain information of this CNV due to the limited size of individual reads. We adopted long read low coverage sequencing that facilitated a method to observe the CNV and generate reliable genotypes to highlight its influence on TSWV. This study demonstrates the utility of population-specific pangenomics and their potential to improve our understanding of intricate traits for crop improvement.

Main

Peanut (*Arachis hypogaea* L.) is produced as an affordable, nutrient rich, and sustainable source of oil and protein worldwide. Because of the economic importance of peanuts as a whole food product and ingredient for many commodities such as peanut oil, peanut butter and candies, there is a constant interest in improving the peanut germplasm in terms of seed quality, resistance to stresses and commercial productivity. *Tomato spotted wilt virus* (TSWV) threatens the existence of peanut cultivation, which is primarily managed through host-resistance with a required integrated pest management system. Therefore, TSWV resistance has been a long-

sought-after trait for preserving genetic gain, facilitating improved resources for TSWV management and production (Agarwal et al., 2019; Khera et al., 2016; Pandey et al., 2017; Qin et al., 2012; Tseng et al., 2016; Wu et al., 2025). The two major challenges of previous QTL studies are marker resolution and recombination frequencies. To overcome these challenges that many plant communities face, the paradigm shift to graph-based pangenome references empowers marker detection to identify recombination patterns more efficiently (Danilevicz et al., 2020; L. Guo et al., 2025; Hou et al., 2024; Liu et al., 2024; Thompson et al., 2025; Vaughn et al., 2022). A population-specific pangenome offers an unbiased reference that integrates a comprehensive set of segregating variants within a population to improve association studies(Thompson et al., 2025; Vaughn et al., 2022) .

The peanut multiparental advanced generation intercross (PeanutMAGIC) population has been previously described as a balanced, diverse population to identify traits of interest through increased recombination events in distinct genetic backgrounds (Thompson et al., 2024, 2025). The PeanutMAGIC founder genomes ‘Georgia-13M’ (13M) (Branch, 2014), ‘SunOleic 97R’ (97R) (Gorbet & Knauft, 2000), ‘GT-C20’ (C20) (Liang et al., 2005), ‘Florida-07’ (F07) (Gorbet & Tillman, 2009), ‘NC94022’ (NC) (Culbreath et al., 2005), ‘TifNV-High O/L’ (TNV) (Holbrook et al., 2017), ‘Tifrunner’ (TR) (Holbrook & Culbreath, 2007), and ‘GP-NC WS16’ (WS16) (Tallury et al., 2014) have been sequenced and used to construct the PeanutMAGIC population-specific pangenome that was successfully leveraged to identify a novel FAD2 gene to resolve the long standing mid-oleic fatty acid mystery (Thompson et al., 2025). Among the founders of PeanutMAGIC, NC is known for TSWV resistance (Culbreath et al., 2005), however it has limited use in commercial settings and breeding programs due to poor seed traits and growth habit. While the data from PeanutMAGIC is proving exceptionally powerful, the

discernment of exceedingly complex phenotypes such as pathogen resistance has not been assessed. Therefore, we hypothesize that PeanutMAGIC and population-specific pangenome are integral for the discernment of the functional variant that confers TSWV resistance to facilitate an improved and accessible germplasm.

In this study, we utilized the MAGIC Core (310 subset of PeanutMAGIC) and 'S' ('NC94022' (NC) X 'SunOleic 97R' (97R)) populations to isolate the TSWV resistance locus and identified a putative causal variation that can be utilized in breeding programs to generate superior resistant lines for growers to integrate into their integrated pest management system. Here we identified a copy number variant that stems from the resistant founder NC in both the MAGIC Core and 'S' populations, containing tandem copies of a novel glutamate receptor. These genes appear to offer a quantitative resistance that segregates as a qualitative trait, simplifying breeding strategies and improving resistance longevity. The findings of population-specific pangenomics improve our understanding of peanut resistance to TSWV and facilitate its application into breeding programs.

Results

Pangenome-based Markers Empower TSWV Resistance Mapping

The MAGIC Core is a subset of the whole PeanutMAGIC population, comprised of 310 $F_{5:7}$ recombinant inbred lines (RILs) that have undergone low coverage (1x) Illumina whole genome sequencing and pangenomic marker calling to yield markers that preserve the founder origin of a locus, while maintaining a consistent coordinate system based on the reference TR (Thompson et al., 2024, 2025). A total of 463,273 markers, including SNP, InDel, and multiallelic sites, were identified. The MAGIC Core was subjected to three years of natural inoculation of TSWV in a randomized complete block design with three replicates per year. The

natural inoculation technique affords the opportunity for variance in inoculum for individual plants and plots. This can create an instance where a susceptible line can escape the virus by receiving little to no inoculum resulting in no apparent symptomology, thus skewing the average phenotype as more resistant. Therefore, to suppress this potential error, the most susceptible phenotypic score among all replicates was used to capture these escapes, thus termed the "Capture" phenotype. Genome wide association study (GWAS) was used to identify regions within the MAGIC Core associated with TSWV resistance. We associated the Capture phenotype with PeanutMAGIC pangenome-based markers and found significant associations on Chr.01 (Fig. 4.1a). Upon examination of Chr.01, there are associations throughout the chromosome that are significant ($P = 2.15x10^{-9}$), the strongest at 12.9 Mb (Fig. 4.1b, orange highlight). To understand the associated signal throughout the chromosome, we examined r^2 values across Chr.01 and found the pericentromere is recalcitrant to recombine within the MAGIC Core in comparison to the telomeric regions (Fig. 4.1c). We further explored the origin of the Chr.01 pericentromere and identified that NC has a unique pericentromere from the other seven founders in PeanutMAGIC with small, infrequent recombinant blocks exchanging and seldom large blocks (Fig. 4.1d; Supplementary Fig. 4.1). The unique variants and suppression of recombination throughout the chromosome may suggest this region has a different origin than the other Chr.01 pericentromeres. By NC having a unique pericentromere and the resistance loci, this makes separating them difficult to consistently achieve for each individual marker and generates signal throughout the chromosome.

In the case of single reference marker calling compared to population-specific pangenome-based markers, we associated the Capture phenotypes with single reference markers for the MAGIC Core (Thompson et al., 2024). Using a Bonferroni-corrected *P* value of 0.001 (*P*

= $7.24x10^{-9}$) as the significant threshold, significant association signals were identified on Chr.01 and Chr.11 (Supplementary Fig. 4.2). Additionally, the most significant association on Chr.01 is at 45 Mb, the middle of the chromosome. Previous findings have identified both the middle and beginning of Chr.01 but have not identified Chr.11 (Supplementary Fig. 4.3; Agarwal et al., 2019; Qin et al., 2012; Tseng et al., 2016; Wu et al., 2025). Three of the four studies support the beginning of Chr.01 for the region where the resistance from NC resides, aligning with the findings of the pangenome-based study (Agarwal et al., 2019; Qin et al., 2012; Wu et al., 2025). The MAGIC Core single reference association has previously demonstrated that it can highlight improper regions on a chromosome that possess the functional variant and identify false variants on a separate chromosome with oleic acid content, due to improper genotype calling using a single reference (Thompson et al., 2025). To investigate if improper genotype calling is causing false association for TSWV resistance, we extracted the associated region on Chr.11, which spans 43 Kb (81,296,099 – 81,339,263) from the PeanutMAGIC pangenome and mapped the 25 markers within the area. We found that none of the 25 single reference markers aligned with a variant in the pangenome, suggesting that these markers are improperly called (Supplementary Fig. 4.4). The false calling of markers can occur due to the inability to index other potential variants of a particular site using a single reference. Single reference markers are inferred based on consistent differences in sequence reads that align to the same location. However, short read sequences can prove difficult to map in repetitive polyploid genomes such as peanut, making accurate marker calling a formidable challenge without the additional information present within a graph-based pangenome. Furthermore, single reference-based markers fail to capture the reduced recombination of the NC pericentromere and overestimate recombination levels due to a lack of information throughout Chr.01, demonstrating the improvements of genomic analysis

with a population-specific pangenome (Supplementary Fig. 4.5). The inability of the single-reference markers to appropriately estimate recombination could be the cause of the middle of Chr.01 being identified using single reference markers and one previous study (Tseng et al., 2016), opposed to the beginning of Chr.01 as found in the pangenome-based markers and three previous studies (Agarwal et al., 2019; Qin et al., 2012; Wu et al., 2025). These findings highlight the importance of a graph-based population-specific pangenome for genomic studies and how they can empower and refine association studies.

To understand the segregation of the causal variant for TSWV resistance in a traditional biparental population, which better represents peanut breeding programs, the 'S' (NC X 97R) population was used for association. The 'S' population has been used to identify TSWV resistance loci in previous studies and has undergone whole genome Illumina sequencing (5x) for 144 RILs (Agarwal et al., 2019; Khera et al., 2016; Qin et al., 2012; Tseng et al., 2016; Wu et al., 2025). Sequence reads were aligned to the 'S' pangenome to generate biparental pangenome-based markers, totaling 113,056 markers (Thompson et al., 2025). To associate a genomic region with TSWV resistance, we followed the Capture phenotyping approach. We identified a location on Chr.01, consistent with the MAGIC Core (Fig. 4.2a). Upon closer examination of Chr.01 it was evident that there is a pinpoint association (12.3 Mb) in the same region as the MAGIC Core (12.9 Mb), however without other signals throughout Chr.01 (Fig. 4.2b). The precise signal from the association would suggest the 'S' population has more recombination than the MAGIC Core, contrary to the rationale of a MAGIC population. To test this hypothesis, we explored r^2 values of Chr.01 in the 'S' population and found that there is considerably more LD in the 'S' population (average r^2: 0.847) than the MAGIC Core (average r^2: 0.438), suggesting recombination is not the only influential factor for the association signal (Fig. 4.2c). We found

that the ‘S’ population can identify the unique NC pericentromere and its recombination pattern, consistent with the MAGIC Core findings (Fig. 4.2d, Supplementary Fig. 4.6).

To elucidate the difference in association signals between the MAGIC Core and ‘S’ population, we examined the qualities of each population structure. PeanutMAGIC RILs have a theoretical 12.5% contribution from the resistant founder NC. In the ‘S’ population the theoretical contribution of NC is 50%. The difference in contribution of the resistant founder within these populations creates a difference in the number of observed lines with and without the NC pericentromere. The observed inheritance percentages of the NC pericentromere was 14.5% in the MAGIC Core (45 RILs) and 50.7% in the ‘S’ population (73 RILs). This indicates that the ‘S’ population had more lines to observe recombination of the NC pericentromere, thus more resistance phenotype to marker comparisons. Of the 45 observed MAGIC Core RILs with the NC pericentromere, 23 (51.11% of the MAGIC Core NC pericentromere subset) are resistant and 22 (48.89% of the MAGIC Core NC pericentromere subset) are susceptible. Only 77 lines of the MAGIC Core need to have information of a site for it to be a marker, based on the 25% minimum representation requirement (Thompson et al., 2025). Therefore, the scenario of a marker that is possessed by resistant lines with the unique NC pericentromere could have limited representation for susceptible lines that harbor the NC pericentromere and still meet the 25% requirement due to the site being detected in RILs that are susceptible and do not possess the NC pericentromere. This allows the marker that is a part of the NC pericentromere, distant from the resistance loci, to be statistically significant, however a false positive. To test this hypothesis, we examined the 15 most significant markers and compared their presence in the 45 PeanutMAGIC RILs. We found that all sites had more than twice the representation in the resistant lines than the susceptible lines (Supplementary Table 4.1). This suggests there was not enough representation

from susceptible lines with the NC pericentromeric region to mediate the resistant line signal. Therefore, when using a multiparental population for isolating a trait of interest, there must be enough recombinant lines that harbor the trait to adequately mitigate linkage of unique genomic features and consistent sequence coverage of sites to fulfil missing data. To address this issue, sequencing of the entire PeanutMAGIC population facilitates the ability to select enough lines that have founder specific genomic features in regions of interest to fine-tune a PeanutMAGIC subpopulation for targeted association studies. Furthermore, enrichment sequencing of targeted sites will ensure ample coverage of sites of interest. In the future, we plan to address these concerns through the availability of the whole PeanutMAGIC population sequencing information, facilitating targeted association of complex traits in a collaborative manner. Together these aspects of materials, resources, and community are integral to the improvement of cultivated peanut production.

Isolation of Resistance Copy Number Variant and Implications in Breeding Populations

The MAGIC Core and 'S' populations identified a similar region on Chr.01 that is consistent with previous studies (Agarwal et al., 2019; Qin et al., 2012; Wu et al., 2025). To pinpoint the resistance locus, we explored the genotyping information of these regions for both populations. We found in the MAGIC Core there is a region from 12.32 - 12.35 Mb that appears necessary for resistance (Fig. 4.3a). In the 'S' population we found a region from 12.32 - 12.36 Mb that is consistent for resistance (Fig. 4.3b). Within the 12.32 - 12.35 Mb region there is a 63.9 Kb structural variant that stems from the founder NC (Fig. 4.3c). The variant harbors part of a kinetochore gene from Chr.20, *arahy.V2489D* and fragments of two neighboring glutamate receptor genes, *arahy.NV21FB* and *arahy.VZ8YXD*. We found that there were four units of sequence that are nearly identical within the insertion. Specifically, there are only three SNPs

between the four units (Supplementary Fig. 4.7). One SNP is an A/G site in the intron of the gene fragment of *arahy.NV21FB* and the other two SNPs sites are between the *arahy.NV21FB* and *arahy.VZ8YXD* gene fragments. The consistency between the four units suggests that they have similar time of origin. The high level of similarity between the native glutamate receptor and the insertion glutamate receptor and its exclusivity to NC, may suggest the introduction of this variant has occurred recently. The identification of a variant within the associated region that consist of four near identical units in tandem succession that is exclusive to the resistant founder NC suggests that the identified copy number variant (CNV) may confer the resistance phenotype of interest.

To observe the potential differences in gene products between the native genes and the gene fragments within the CNV, we adopted recent advances in protein structure prediction. Using *arahy.V2489D* exon annotations for the gene fragments in the CNV, the translated sequence resulted in small open reading frames throughout the coding sequence that aligned to kinetochore proteins and shares a similar protein structure with *arahy.V2489D* (Supplementary Table 4.2 & 4.3, Supplementary Fig. 4.8). The sequence for the fragmented *arahy.VZ8YXD* was only the last exon of the annotated protein, which would likely not be transcribed or translated (Supplementary Fig. 4.9). Monomer prediction of the *arahy.NV21FB* gene fragment showed a truncated protein product, however maintaining the conserved motif of a glutamate receptor and aligns to other annotated glutamate receptors (Fig. 4.3d, Supplementary Tables 4.4-4.6). Therefore, the most likely functional protein of the variant is the truncated *arahy.NV21FB* gene that has been duplicated four times within the copy number insertion.

The CNV is present in both the PeanutMAGIC pangenome and within the ‘S’ pangenome, however marker calls were not made for this location, likely due to its repetitive

nature and short read sequencing data. To genotype the population for this variant, personalized pangenomes were constructed for depth calling of the area (Hickey et al., 2020; Sirén et al., 2021, 2024). This approach facilitates variance calling for the individuals of the MAGIC Core population through manual observation (Fig. 4.4a-4.4c). Low coverage short read sequencing can collect some sequence information of a structural variant or CNV. However, complete alignment throughout these variants is challenging due to multiple alignment with identical sequences. This ambiguity makes consistent mapping of the entire variant challenging, however sites within the variant can be aligned for detection of the variant with some insight into variant completeness (Fig. 4.4a). For lines without the CNV, they may not have any reads mapped to the region (Fig. 4.4b). Alternatively, lines without the CNV can possess identical sequence to the structural variant, which can improperly align to the variant in select places (Fig. 4.4c). Conversely, lines possessing the CNV may have reads of the variant that improperly align to common sequence increasing values elsewhere and limited alignments within the region. We then compared the presence and absence of the insertion to their respective Capture phenotype and found a significant reduction in disease severity with the presence of the CNV (Students *t*-test $1.5x10^{-5}$), supporting that the CNV is influential to TSWV symptomology. We also used personalized pangenomes for the 'S' biparental population. The 'S' population was sequenced with higher depth (5x) than the MAGIC Core (1x), simplifying manual observation of the CNV (Fig. 4.4e-g). Additionally, off target mapping is more prevalent due to more observations of sequence, however the difference between ambiguous mapping and coverage of the CNV are more discernable (Fig. 4.4e-g). We tested the significance of the CNV for disease control and found a significant reduction in TSWV severity with the presence of the CNV within the 'S' biparental population. Together, the MAGIC Core and 'S' population found a significant

reduction in TSWV severity with the presence of the CNV from the resistant founder NC, suggesting that this CNV can improve TSWV management and can be identified within a breeding program.

To highlight the significance of the CNV for TSWV resistance, we used the presence-absence variation to generate a CNV marker for GWAS analysis. In the MAGIC Core population, we found that the CNV marker was significantly associated with the Capture phenotype, although it was not the most significant signal (Fig. 4.4i). When the CNV marker signal was compared to the pangenome-based marker signal, it is evident that the number of observations influences the signal from GWAS. For the CNV marker all 310 lines had a call for either presence or absence. However, the number of calls for the most significant pangenome-based marker totaled 78, due to many of the lines not possessing sequence data to support a variant call for that location. This allows for a marker with less observations to have more significance due to the number of lines with the unique marker expressing resistance. We found 8.06% (25 of 310) of the lines with the CNV were resistant, where 84.62% (66 of 78) of lines with the most significant pangenome-based marker were resistant, consequently overpowering the CNV signal. In a comparison of pangenome-based markers with similar depth throughout the population the discrepancy of marker depth is not present (Supplementary Fig. 4.10), however the CNV marker with complete data became suppressed from limited observations possessing the unique variant. In the case of the ‘S’ biparental population, the CNV marker had the highest association signal (Fig. 4.4j). This is due to the similar representation of the highest *P*-value pangenome-based marker compared to the CNV marker (53.68% and 50.69% of marker representation from resistant lines, respectively). The equal representation ratio of the resistant founder and the increased sequence depth for the ’S’ population (5x compared to MAGIC Core

1x), yielded more marker depth than in the MAGIC Core and NC specific marker observations (Supplementary Fig. 4.11). In a biparental population the percent of the population that possesses the variant to produce an adequate signal aligns with Mendelian segregation, however in a multiparental population this is not the case. When a single founder harbors the resistance a small portion of a random selection will possess the resistance and can be in linkage to areas with increased diversity, convoluting association analysis. Thus, sequencing the whole population and selecting lines with equal representation of unique regions should improve association signal. Furthermore, signals from enriched sites of interest can be overwhelmed by signals of unique neighboring sites with less depth throughout the population. This suggests enriched sites should be understood differently than whole genome sequencing sites during data analysis, particularly in a multiparental population framework.

Upon further investigation of the CNV identification method, we found that short read sequencing captures only a fraction of the insertion and these regions could be misaligned due to the limited information retained within each read and the size of the read compared to the CNV (Supplementary Fig. 4.12). Therefore, we adopted long read sequencing techniques to cover the CNV to overcome the limitations of short read sequencing. We selected 94 MAGIC Core lines that were subject to low coverage long read sequencing to generate sequence reads that would span into the CNV for improved genotyping. To understand the differences between short and long read information of the CNV we plotted the reads to scale of the CNV and compared genotyping consistency. MG1104 is a line that is susceptible to TSWV and was identified to not possess the CNV in long and short read sequences (Fig. 4.5a & 4.5b). In a similar fashion, MG1304 is a resistant line and was found to possess the CNV in both the short and long read data sets (Fig. 4.5c & 4.5d). However, it became evident that the short read data can produce

errors in CNV genotyping. We found in susceptible line MG206, short read data mapped to four locations within the CNV and was called present (Fig 4.5e). Using long read data it became evident that MG206 does not possess the CNV, demonstrating that short read data can produce false positives for CNV calling (Fig. 4.5f). Inversely, the resistant line MG209 short read data failed to have reads map within the CNV, thus was called absent (Fig. 4.5g). However, long read low coverage data of MG209 mapped throughout the CNV, demonstrating that short read data can also produce false negatives for CNV calling (Fig. 4.5h). To understand the discrepancies between short and long read sequences we compared the coverage of the CNV in lines that possessed the variant in the long read data. The coverage of the CNV was calculated by taking the area which a read aligned making it independent of the sequence depth of the CNV. We observed significant differences between the short read and long read coverage of the CNV (Fig. 4.5i). This difference in the ability of the data set to retain information of the CNV highlights the difficulties of short read sequencing to investigate large, repetitive variants that can drive plant diversity and important agricultural phenotypes. We tested the significance of the CNV to the Capture phenotype for the long read lines and identified a higher significance value compared to the short read data ($1.20x10^{-6}$ and $1.50x10^{-5}$, respectively; Fig. 4.5j) although a smaller sample size. The use of long read low coverage sequence data allowed for clear observations of the CNV for genotyping. Short read sequences contain a fraction of the information needed to understand a CNV and continue to produce ambiguous mapping signal even with additional coverage depth as in the ‘S’ population. Long read sequences can span several Kb facilitating their mapping and interpretation clear for genotype calling of challenging putative functional variants such as CNV in addition to other variants.

A potential mechanism of resistance of the CNV is the increased copy number of

glutamate receptors that improves cellular signaling to the inoculation of TSWV, stimulating an improved defense response to yield higher levels of tolerance. In tomatoes and pepper, TSWV resistance is controlled by a single dominant gene (*Sw5* and *Tsw*, respectively) that initiates a hypersensitivity response to confer resistance (Leastro et al., 2017; Moury et al., 1997; Spassova et al., 2001). To distinguish the differences between the types of resistance in peanut and in *Solanaceae*, observation of phenotypes collected show that the resistance in PeanutMAGIC can be overcome at points but are maintained overtime (Supplementary Table 4.7). Once a hypersensitivity response has been broken, the virulent strains can freely spread to other resistant lines and persist over time. In the MAGIC Core, lines can show increased symptomology for independent plots but not throughout the field or in later seasons, indicating that the resistance within peanut has not been broken but simply overloaded. This theory explains the increased values of Capture phenotypes with lines possessing the insertion due to random over inoculation for these individuals. Glutamate receptors are known for signal cascades for herbivory damage and pathogen presence, which are necessary for host resistance, however, can have limitations with severe disease pressure (F. Li et al., 2013; Nguyen et al., 2018; Sun et al., 2019). Additionally, CNV can facilitate a manner for a single low effect gene to function as a quantitative trait, however segregates as a qualitative trait, simplifying breeding selection and minimizing the potential of the virus overcoming the resistance. Together these findings support the theory that the resistance mechanism is through rapid activation of generalized host defense, that can be overloaded but is reliable overtime.

Discussion

Peanuts are an affordable, nutrient rich, and sustainable crop throughout the world that supply food and ingredients for human consumption. The identification of genetic markers that

accurately track genomic sequences is a requirement for genomic studies and trait associations. Plant genomics that rely on a single reference linear genome have the potential to introduce reference bias through a lack of information of other possible variants. This study utilizes population-specific pangenomes that represent a paradigm shift from a single species reference. By including all founders of a synthetic population into one resource, a population-specific pangenome is a comprehensive library of potential segregant variants that can be used as genomic markers for analysis and association.

To demonstrate the power of population-specific pangenome-based associations and analysis, we focused on the required agronomic trait TSWV resistance and the elusive variant from NC to improve cultivar resistance. The increased information available through pangenome-based markers allowed for better understanding of traditional recombination cold-spots to precisely identify the putative causal variation for TSWV resistance. We identified a CNV containing a novel glutamate receptor that may function as a quantitative basal defense activator. The CNV was found to significantly reduce the most severe disease rating a genotype possessed in both the MAGIC Core and 'S' populations. This finding underscores the improvements of association mapping studies when utilizing pangenome-based genotyping through resolving a long-standing question within the peanut community.

In the case of application in breeding programs, the CNV possesses multiple genes that are observed as a quantitative phenotype, but the genes are linked together facilitating their inheritance pattern as a qualitative trait. This offers a trait that should yield long term control for TSWV in parallel to other quantitative resistance. Importantly, these genes are linked giving the added benefit of their connected segregation, simplifying marker assisted selection and ease of pyramiding other important agronomical traits. Furthermore, the hypothesized resistance

mechanisms may not require TSWV or peanut specific interactions giving the potential for this system to improve resistance to other peanut pathogens such as groundnut rosette virus or the CNV transformed into other systems such as tomato to subside qualitative resistance breakdown. Population-specific pangenomics empowers genomic analysis and association studies for an unbiased approach that will improve basic research and cultivar breeding efforts.

Methods

Plant Materials and Short Read Sequencing

The PeanutMAGIC population has been described previously (Thompson et al., 2024). Briefly, eight founder lines Georgia-13M (13M), SunOleic 97R (97R), GT-C20 (C20), Florida-07 (F07), NC94022 (NC), TifNV-High O/L (TNV), Tifrunner (TR), and GP-NC WS16 (WS16) were crossed in a simple funnel-like design. The critical 8-way cross consisted of 150 successful pairs, generating 950 unique F_1 offspring. The population was advanced to form 3,187 $F_{2:7}$ RILs. A random selection of 310 RILs were collected for this study termed the MAGIC Core. RILs were subject to CTAB DNA extraction and low-coverage (~1x) Illumina sequencing (Thompson et al., 2024).

The ‘S’ population has previously been described (Agarwal et al., 2019; Khera et al., 2016). Briefly, two lines, SunOleic 97R and NC94022, were selected to develop a RIL mapping population consisting of 352 RILs. The two founders of the ‘S’ population are founders of the PeanutMAGIC population. Of the whole ‘S’ population, 144 RILs were selected and previously sequenced with 5x Illumina sequencing (Agarwal et al., 2019) and were utilized in this study.

Founder Assembly and Pangenomes Construction

The founders of the PeanutMAGIC population have been sequenced previously using long read sequencing approaches (Thompson et al., 2025; (Bertioli et al., 2019). A population-

specific pangenome for PeanutMAGIC has been constructed previously using the Mini-Graph Cactus Pangenome Pipeline (v2.6.7) (Armstrong et al., 2020; Hickey et al., 2023) with TR as the reference genome (Thompson et al., 2025). To generate a population-specific pangenome for the 'S' population, the 97R and NC assemblies were unified using Mini-Graph Cactus Pangenome Pipeline where NC was the reference (Thompson et al., 2025). To assess the graphs vg (v1.56.0) (Garrison et al., 2018) subcommands 'stats' and 'deconstruct' (Liao et al., 2023) were used to extract components and variant locations for each pangenome.

Khufu Pangenomic Marker Calling

KhufuPAN has been described in part (Thompson et al., 2025). Briefly, KhufuPAN requires a graphical fragment assembly (GFA) file. The file should include the genomes forming the pangenome and a well-assembled reference genome (designated null reference), which is used to assign the coordinates of SNPs and structural variants. The GFA is deconstructed to produce a parental VCF file. Then a series of filters is applied to remove odd alleles, low-quality variants, monomorphic variants, or those having polymorphism only with the null reference genome to generate a Filtered-Variant set. KhufuVAR is a KhufuPAN sub-tool used to call and filter GFA variants for markers. Raw FASTQ files, short or long-reads, are applied to Khufu-core (https://www.hudsonalpha.org/khufudata/), and then mapped to the GFA file using vg giraffe (https://github.com/vgteam/vg/wiki/Mapping-short-reads-with-Giraffe) (Sirén et al., 2021). Calls under the minimum depth cutoff are masked, and variants overlapping with the Filtered-Variant set are extracted for genomic analysis. The final output calls are exported in Khufu panmap format, which facilitates pangenome-based markers for genomic analysis. The generated panmap file can then be used to extract a rds file for Cyclops to visualize markers throughout the population in chromosomal units (https://w-korani.shinyapps.io/cyclops_eye_ii/).

TSWV phenotyping and GWAS Analysis

To collect phenotypes for association, plots of two rows 1.2m long, containing 20 seeds were planted in Tifton, GA over three growing seasons for the MAGIC Core and the 'S' populations. Each season three replicates were planted in a randomized block design and were naturally inoculated. TSWV phenotypes are scored on a 1-5 scale, where 1 represents no symptomology and 5 represents sever stunting, chlorosis, and spotting. For each plot the most severe plant is used to represent the plot. The "Capture" phenotype is the most severe score observed throughout all replicates and seasons.

Pangenome markers were used to calculate a distance matrix for multidimensional scaling to control population structure for MAGIC Core pangenome-based markers, MAGIC Core single reference-based markers and 'S' pangenome-based markers. A general linear model was adopted for GWAS analysis on TASSEL 5.0 software (Bradbury et al., 2007). TSWV Capture phenotypes and genotypes were used in conjunction with the model to isolate regions of interest. This process was used for MAGIC Core pangenome-based markers, MAGIC Core single reference-based markers, and 'S' pangenome-based markers with respective phenotypes and genotypes for consistent parameters between associations.

MAGIC Core Pangenome Based Genomic Characterization

To understand marker segregation within the MAGIC Core population, r^2 values were calculated in a 300-marker sliding window on TASSEL 5.0 software (Bradbury et al., 2007) and plotted by the physical location. The number of 300 markers was chosen based on the increase in marker density compared to the sliding window used in Thompson et al., 2024. To maintain similar physical window sizes the number of markers per window was increased according to the average increase in markers (3x), simplifying single and pangenome reference comparison.

In silico protein prediction

Protein structures of *arahy.V2489D, arahy.NV21FB, arahy.VZ8YXD,* and gene fragments were computationally predicted using Alphafold (v2.3.1) (Jumper et al., 2021). Annotated protein sequences were used to generate monomers for each annotated gene and annotated exon fragment alignments were translated for gene fragments. The databases used for structure construction were BFD database "bfd_metaclust_clu_complete_id30_c90_final_seq.sorted_opt", the 'Mgnify' database from May 2022, the 'pdb70' database from March 2022, the 'Uniclust 30' database from March 2021, and the 'Uniref 30' database from March 2021.

Personalized Pangenome Genotyping

Sequencing reads for MAGIC Core were used to generate personalized pangenomes for each individual following the methods on the vg GitHub wiki (https://github.com/vgteam/vg/wiki/Haplotype-Sampling) (Sirén et al., 2024). First, this method uses kmc to count *k*-mers from the sequence reads (Kokot et al., 2017). Next, the *k*-mer counts are implemented into haplotype sampling using vg giraffe (Sirén et al., 2021). The subsequent GBZ and minimizer files are then used with vg 'pack' to compute read support for genotyping (Garrison et al., 2018). The genotyping calls were then made using vg 'call' (Hickey et al., 2020). This workflow generated personalized pangenomes and genotypes that were used to extract sequence depth information for the insertion across the MAGIC Core and the 'S' populations.

Statistical Testing

All phenotypic statistical analysis were conducted and visualized using R statistical software (R Core Team, 2014, 2021) with the package ggplot2 (Wickham, 2011), respectively. Student's *t* test was performed using ggpubr (Kassambara, 2020) function

'stat_compare_means(method = "t.test")'.

Whole Genome Long Read Low Coverage Sequencing

High molecular weight DNA was extracted from 94 MAGIC Core lines using the unofficial Pacbio high-throughput gDNA workflow for plants (https://www.pacb.com/support/documentation/). The subsequent DNA was then multiplexed and formed libraries using Pacbio Hifi plex prep kit for Revio sequencing. The reads were then separated into their respective genotypes to generate raw long reads. The long reads were mapped to the PeanutMAGIC pangenome using the personalized pangenome genotyping approach and verified using minimap (H. Li, 2018, 2021).

References

Agarwal, G., Clevenger, J., Kale, S. M., Wang, H., Pandey, M. K., Choudhary, D., Yuan, M., Wang, X., Culbreath, A. K., Holbrook, C. C., Liu, X., Varshney, R. K., & Guo, B. (2019). A recombination bin-map identified a major QTL for resistance to Tomato Spotted Wilt Virus in peanut (*Arachis hypogaea*). *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-54747-1

Armstrong, J., Hickey, G., Diekhans, M., Fiddes, I. T., Novak, A. M., Deran, A., Fang, Q., Xie, D., Feng, S., Stiller, J., Genereux, D., Johnson, J., Marinescu, V. D., Alföldi, J., Harris, R. S., Lindblad-Toh, K., Haussler, D., Karlsson, E., Jarvis, E. D., … Paten, B. (2020). Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature*, *587*(7833). https://doi.org/10.1038/s41586-020-2871-y

Bertioli, D. J., Jenkins, J., Clevenger, J., Dudchenko, O., Gao, D., Seijo, G., Leal-Bertioli, S. C. M., Ren, L., Farmer, A. D., Pandey, M. K., Samoluk, S. S., Abernathy, B., Agarwal, G., Ballén-Taborda, C., Cameron, C., Campbell, J., Chavarro, C., Chitikineni, A., Chu, Y., …

Schmutz, J. (2019). The genome sequence of segmental allotetraploid peanut Arachis hypogaea. *Nature Genetics*, *51*(5). https://doi.org/10.1038/s41588-019-0405-z

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). https://doi.org/10.1093/bioinformatics/btm308

Branch, W. D. (2014). Registration of 'Georgia-13M' Peanut. *Journal of Plant Registrations*, *8*(3). https://doi.org/10.3198/jpr2013.11.0071crc

Culbreath, A. K., Gorbet, D. W., Martinez-Ochoa, N., Holbrook, C. C., Todd, J. W., Isleib, T. G., & Tillman, B. (2005). High Levels of Field Resistance to Tomato spotted wilt virus in Peanut Breeding Lines Derived from hypogaea and hirsuta Botanical Varieties. *Peanut Science*, *32*(1). https://doi.org/10.3146/0095-3679(2005)32[20:hlofrt]2.0.co;2

Danilevicz, M. F., Tay Fernandez, C. G., Marsh, J. I., Bayer, P. E., & Edwards, D. (2020). Plant pangenomics: approaches, applications and advancements. In *Current Opinion in Plant Biology* (Vol. 54). https://doi.org/10.1016/j.pbi.2019.12.005

Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*. https://doi.org/10.2307/4119796

Garrison, E., Sirén, J., Novak, A. M., Hickey, G., Eizenga, J. M., Dawson, E. T., Jones, W., Garg, S., Markello, C., Lin, M. F., Paten, B., & Durbin, R. (2018). Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nature Biotechnology*, *36*(9). https://doi.org/10.1038/nbt.4227

Gorbet, D. W., & Knauft, D. A. (2000). Registration of 'SunOleic 97R' Peanut. *Crop Science*, *40*(4). https://doi.org/10.2135/cropsci2000.0032rcv

Gorbet, D. W., & Tillman, B. L. (2009). Registration of ‘Florida-07’ Peanut. *Journal of Plant Registrations*, *3*(1). https://doi.org/10.3198/jpr2008.05.0276crc

Guo, L., Wang, X., Ayhan, D. H., Rhaman, M. S., Yan, M., Jiang, J., Wang, D., Zheng, W., Mei, J., Ji, W., Jiao, J., Chen, S., Sun, J., Yi, S., Meng, D., Wang, J., Bhuiyan, M. N., Qin, G., Guo, L., … Ye, W. (2025). Super pangenome of Vitis empowers identification of downy mildew resistance genes for grapevine improvement. *Nature Genetics*, *57*(3), 741–753. https://doi.org/10.1038/s41588-025-02111-7

Hickey, G., Heller, D., Monlong, J., Sibbesen, J. A., Sirén, J., Eizenga, J., Dawson, E. T., Garrison, E., Novak, A. M., & Paten, B. (2020). Genotyping structural variants in pangenome graphs using the vg toolkit. *Genome Biology*, *21*(1). https://doi.org/10.1186/s13059-020-1941-7

Hickey, G., Monlong, J., Ebler, J., Novak, A. M., Eizenga, J. M., Gao, Y., Abel, H. J., Antonacci-Fulton, L. L., Asri, M., Baid, G., Baker, C. A., Belyaeva, A., Billis, K., Bourque, G., Buonaiuto, S., Carroll, A., Chaisson, M. J. P., Chang, P. C., Chang, X. H., … Paten, B. (2023). Pangenome graph construction from genome alignments with Minigraph-Cactus. *Nature Biotechnology*. https://doi.org/10.1038/s41587-023-01793-w

Holbrook, C. C., & Culbreath, A. K. (2007). Registration of ‘Tifrunner’ Peanut. *Journal of Plant Registrations*, *1*(2), 124–124. https://doi.org/10.3198/jpr2006.09.0575crc

Holbrook, C. C., Ozias-Akins, P., Chu, Y., Culbreath, A. K., Kvien, C. K., & Brenneman, T. B. (2017). Registration of ‘TifNV-High O/L’ Peanut. *Journal of Plant Registrations*, *11*(3). https://doi.org/10.3198/jpr2016.10.0059crc

Hou, Y., Gan, J., Fan, Z., Sun, L., Garg, V., Wang, Y., Li, S., Bao, P., Cao, B., Varshney, R. K., & Zhao, H. (2024). Haplotype-based pangenomes reveal genetic variations and climate

adaptations in moso bamboo populations. *Nature Communications*, *15*(1), 8085. https://doi.org/10.1038/s41467-024-52376-5

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., … Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873). https://doi.org/10.1038/s41586-021-03819-2

Kassambara, A. (2020). ggpubr: "ggplot2" based publication ready plots. R package version 0.2. *Https://CRAN.R-Project.Org/Package=ggpubr*.

Khera, P., Pandey, M. K., Wang, H., Feng, S., Qiao, L., Culbreath, A. K., Kale, S., Wang, J., Holbrook, C. C., Zhuang, W., Varshney, R. K., & Guo, B. (2016). Mapping Quantitative Trait Loci of Resistance to Tomato Spotted Wilt Virus and Leaf Spots in a Recombinant Inbred Line Population of Peanut (Arachis hypogaea L.) from SunOleic 97R and NC94022. *PLOS ONE*, *11*(7), e0158452. https://doi.org/10.1371/journal.pone.0158452

Kokot, M., Dlugosz, M., & Deorowicz, S. (2017). KMC 3: counting and manipulating k-mer statistics. *Bioinformatics (Oxford, England)*, *33*(17). https://doi.org/10.1093/bioinformatics/btx304

Leastro, M. O., De Oliveira, A. S., Pallás, V., Sánchez-Navarro, J. A., Kormelink, R., & Resende, R. O. (2017). The NSm proteins of phylogenetically related tospoviruses trigger Sw-5b–mediated resistance dissociated of their cell-to-cell movement function. *Virus Research*, *240*, 25–34. https://doi.org/10.1016/j.virusres.2017.07.019

Li, F., Wang, J., Ma, C., Zhao, Y., Wang, Y., Hasi, A., & Qi, Z. (2013). Glutamate Receptor-Like Channel3.3 Is Involved in Mediating Glutathione-Triggered Cytosolic Calcium

Transients, Transcriptional Changes, and Innate Immunity Responses in Arabidopsis. *PLANT PHYSIOLOGY*, *162*(3), 1497–1509. https://doi.org/10.1104/pp.113.217208

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H. (2021). New strategies to improve minimap2 alignment accuracy. *Bioinformatics*, *37*(23). https://doi.org/10.1093/bioinformatics/btab705

Liang, X. Q., Holbrook, C. C., Lynch, R. E., & Guo, B. Z. (2005). β-1,3-Glucanase Activity in Peanut Seed (*Arachis hypogaea*) is Induced by Inoculation with *Aspergillus flavus* and Copurifies with a Conglutin-Like Protein. *Phytopathology®*, *95*(5), 506–511. https://doi.org/10.1094/PHYTO-95-0506

Liao, W. W., Asri, M., Ebler, J., Doerr, D., Haukness, M., Hickey, G., Lu, S., Lucas, J. K., Monlong, J., Abel, H. J., Buonaiuto, S., Chang, X. H., Cheng, H., Chu, J., Colonna, V., Eizenga, J. M., Feng, X., Fischer, C., Fulton, R. S., … Paten, B. (2023). A draft human pangenome reference. *Nature*, *617*(7960). https://doi.org/10.1038/s41586-023-05896-x

Liu, Z., Wang, N., Su, Y., Long, Q., Peng, Y., Shangguan, L., Zhang, F., Cao, S., Wang, X., Ge, M., Xue, H., Ma, Z., Liu, W., Xu, X., Li, C., Cao, X., Ahmad, B., Su, X., Liu, Y., … Zhou, Y. (2024). Grapevine pangenome facilitates trait genetics and genomic breeding. *Nature Genetics*, *56*(12), 2804–2814. https://doi.org/10.1038/s41588-024-01967-5

Moury, B., Palloix, A., Gebre Selassie, K., & Marchoux, G. (1997). Hypersensitive resistance to tomato spotted wilt virus in three Capsicum chinense accessions is controlled by a single gene and is overcome by virulent strains. *Euphytica*, *94*(1), 45–52. https://doi.org/10.1023/A:1002997522379

Nguyen, C. T., Kurenda, A., Stolz, S., Chételat, A., & Farmer, E. E. (2018). Identification of cell populations necessary for leaf-to-leaf electrical signaling in a wounded plant. *Proceedings of the National Academy of Sciences*, *115*(40), 10178–10183. https://doi.org/10.1073/pnas.1807049115

Pandey, M. K., Wang, H., Khera, P., Vishwakarma, M. K., Kale, S. M., Culbreath, A. K., Holbrook, C. C., Wang, X., Varshney, R. K., & Guo, B. (2017). Genetic Dissection of Novel QTLs for Resistance to Leaf Spots and Tomato Spotted Wilt Virus in Peanut (Arachis hypogaea L.). *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.00025

Qin, H., Feng, S., Chen, C., Guo, Y., Knapp, S., Culbreath, A., He, G., Wang, M. L., Zhang, X., Holbrook, C. C., Ozias-Akins, P., & Guo, B. (2012). An integrated genetic linkage map of cultivated peanut (Arachis hypogaea L.) constructed from two RIL populations. *Theoretical and Applied Genetics*, *124*(4). https://doi.org/10.1007/s00122-011-1737-y

R Core Team. (2014). R Core Team (2014). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria. URL Http://Www.R-Project.Org/*.

R Core Team. (2021). R core team (2021). In *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www. R-project. org*.

Sirén, J., Eskandar, P., Ungaro, M. T., Hickey, G., Eizenga, J. M., Novak, A. M., Chang, X., Chang, P.-C., Kolmogorov, M., Carroll, A., Monlong, J., & Paten, B. (2024). Personalized pangenome references. *Nature Methods*, *21*(11), 2017–2023. https://doi.org/10.1038/s41592-024-02407-2

Sirén, J., Monlong, J., Chang, X., Novak, A. M., Eizenga, J. M., Markello, C., Sibbesen, J. A., Hickey, G., Chang, P.-C., Carroll, A., Gupta, N., Gabriel, S., Blackwell, T. W., Ratan, A., Taylor, K. D., Rich, S. S., Rotter, J. I., Haussler, D., Garrison, E., & Paten, B. (2021). Pangenomics enables genotyping of known structural variants in 5202 diverse genomes. *Science*, *374*(6574). https://doi.org/10.1126/science.abg8871

Spassova, M. I., Prins, T. W., Folkertsma, R. T., Klein-Lankhorst, R. M., Hille, J., Goldbach, R. W., & Prins, M. (2001). The tomato gene Sw5 is a member of the coiled coil, nucleotide binding, leucine-rich repeat class of plant resistance genes and confers resistance to TSWV in tobacco. *Molecular Breeding*, *7*(2), 151–161. https://doi.org/10.1023/A:1011363119763

Sun, C., Jin, L., Cai, Y., Huang, Y., Zheng, X., & Yu, T. (2019). l-Glutamate treatment enhances disease resistance of tomato fruit by inducing the expression of glutamate receptors and the accumulation of amino acids. *Food Chemistry*, *293*, 263–270. https://doi.org/10.1016/j.foodchem.2019.04.113

Tallury, S. P., Isleib, T. G., Copeland, S. C., Rosas-Anderson, P., Balota, M., Singh, D., & Stalker, H. T. (2014). Registration of Two Multiple Disease-Resistant Peanut Germplasm Lines Derived from *Arachis cardenasii* Krapov. & W.C. Gregory, GKP 10017. *Journal of Plant Registrations*, *8*(1), 86–89. https://doi.org/10.3198/jpr2013.04.0017crg

Thompson, E., Korani, W., Wu, D., Garg, V., Tonnis, B., Wang, M., Holbrook, C. C., Ozias-Akins, P., Culbreath, A. K., Varshney, R. K., Guo, B., & Clevenger, J. (2025). PeanutMAGIC multiparent population-specific pangenome unveils a novel third FAD2 gene and solves the mystery of mid-oleic fatty acid in peanut. *Nature Communications, In Review*.

Thompson, E., Wang, H., Korani, W., Fountain, J. C., Culbreath, A. K., Holbrook, C. C., Clevenger, J. P., & Guo, B. (2024). Genetic and genomic characterization of a multiparent advanced generation intercross (MAGIC) population of peanut (*Arachis hypogaea* L.). *Crop Science*, *65*(1). https://doi.org/10.1002/csc2.21402

Tseng, Y.-C., Tillman, B. L., Peng, Z., & Wang, J. (2016). Identification of major QTLs underlying tomato spotted wilt virus resistance in peanut cultivar Florida-EPTM '113.' *BMC Genetics*, *17*(1), 128. https://doi.org/10.1186/s12863-016-0435-9

Vaughn, J. N., Branham, S. E., Abernathy, B., Hulse-Kemp, A. M., Rivers, A. R., Levi, A., & Wechter, W. P. (2022). Graph-based pangenomics maximizes genotyping density and reveals structural impacts on fungal resistance in melon. *Nature Communications*, *13*(1), 7897. https://doi.org/10.1038/s41467-022-35621-7

Wickham, H. (2011). Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. In *Wiley Interdisciplinary Reviews: Computational Statistics* (Vol. 3, Issue 2).

Wu, D., Zhao, C., Korani, W., Thompson, E. A., Wang, H., Agarwal, G., Fountain, J. C., Culbreath, A., Holbrook, C. C., Wang, X., Clevenger, J. P., & Guo, B. (2025). High-resolution genetic and physical mapping reveals a peanut spotted wilt disease resistance locus, PSWDR-1, to Tomato spotted wilt virus (TSWV), within a recombination cold-spot on chromosome A01. *BMC Genomics*, *26*(1), 224. https://doi.org/10.1186/s12864-025-11366-7

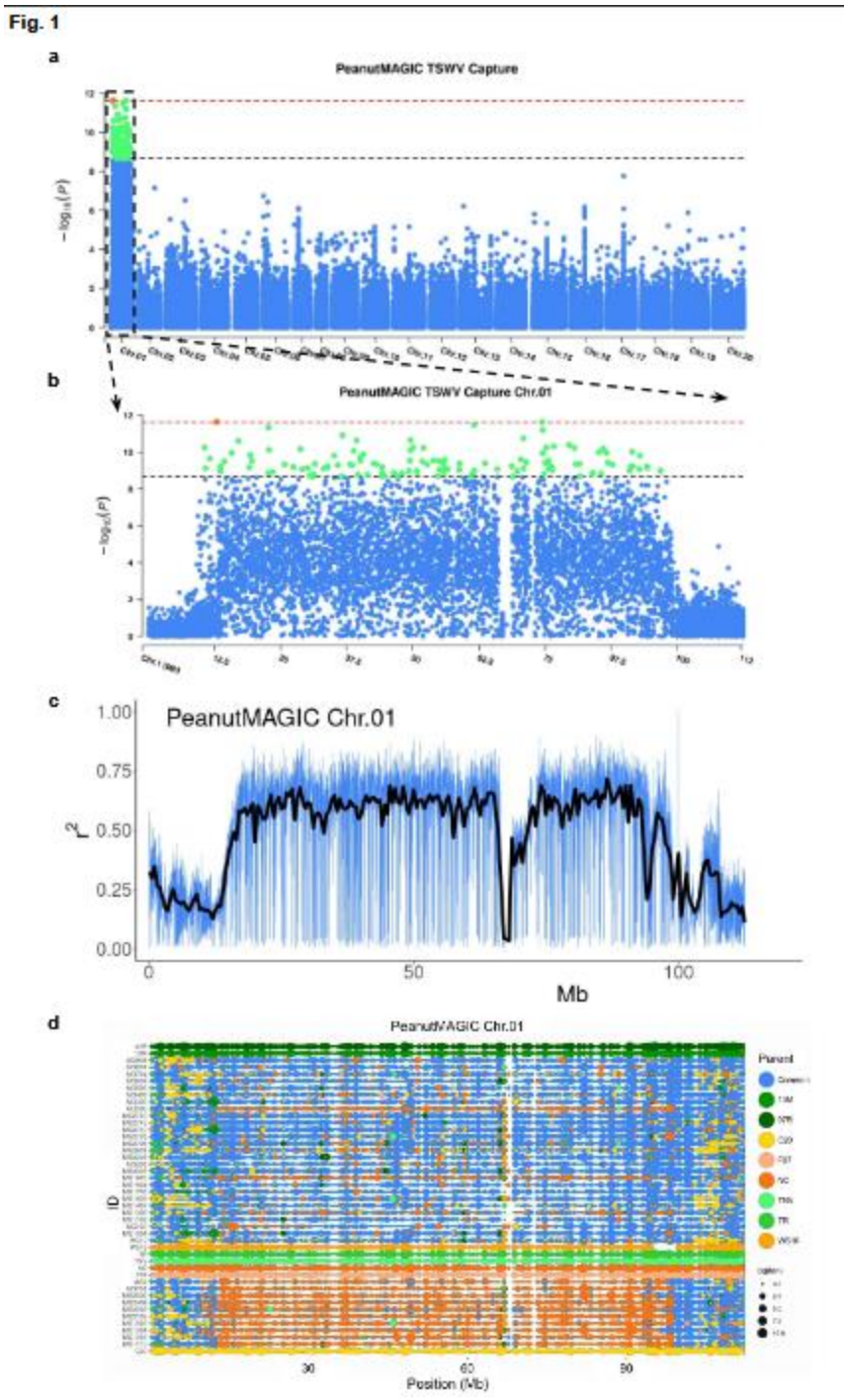


Fig. 4.1: **PeanutMAGIC pangenomic association with Tomato Spotted Wilt Virus (TSWV)**.

a, Manhattan plot showing PeanutMAGIC pangenome based markers associated with TSWV Capture phenotypes. Capture represents the most susceptible phenotype observed

throughout all replicates and seasons. A Bonferroni-corrected *P* value of 0.001 was used as a significant threshold ($P = 2.15x10^{-9}$), represented by a horizontal dashed black line. The red dash line represents a threshold for the highest signal at $P = 2.3x10^{-12}$. **b**, Local association plot of Chr.01 showing significant signals throughout the chromosome. **c**, PeanutMAGIC pangenome marker r^2 values of Chr.01 showing increased linkage disequilibrium throughout the pericentromeric region. **d**, Subset of MAGIC Core RILs with pangenome marker calls for Chr.01 with founder origin coloration. Orange represents markers from the TSWV resistant founder 'NC94022' (NC).

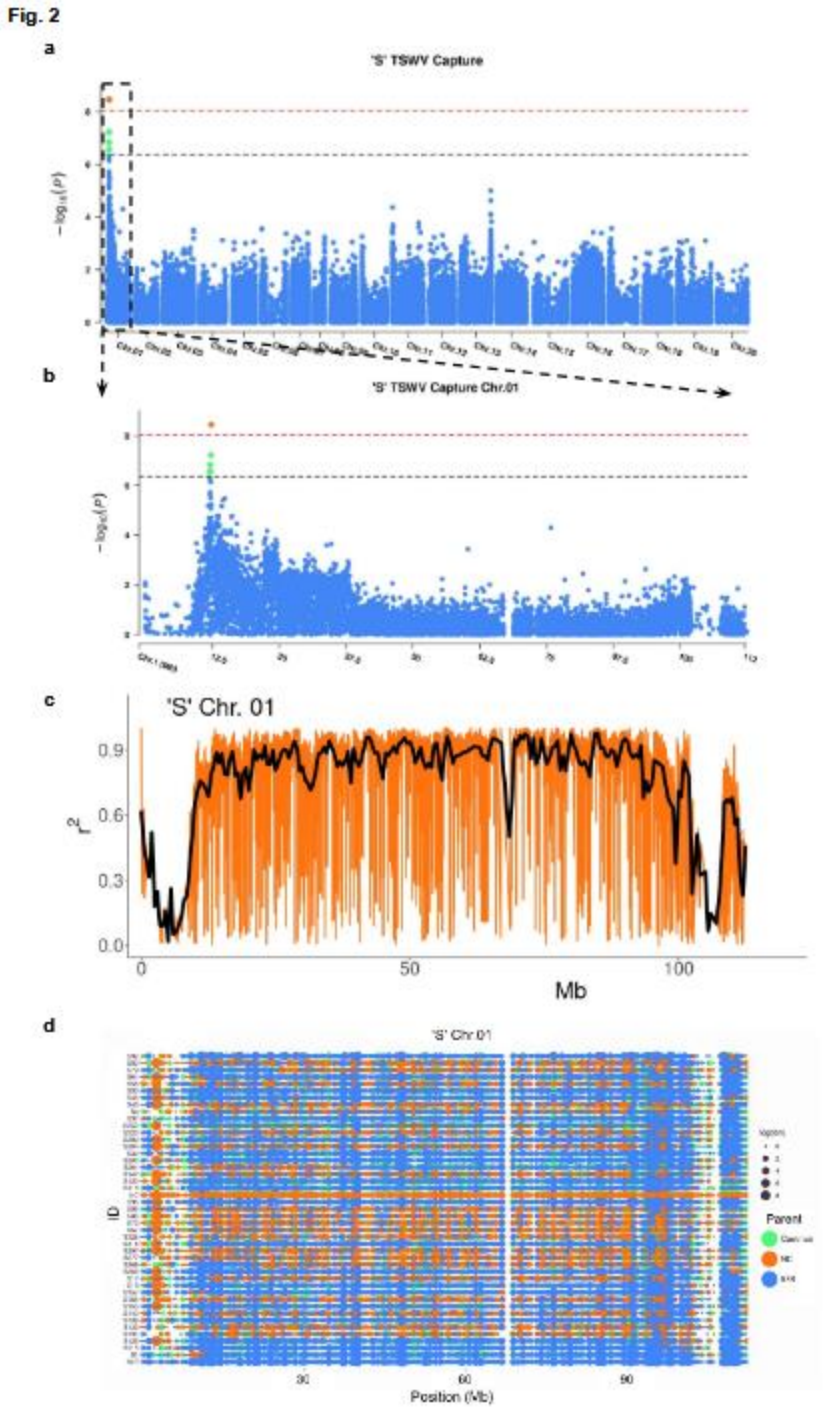


Fig. 4.2: **Biparental ‘S’ pangenomic association with Tomato Spotted Wilt Virus (TSWV)**.

a, Manhattan plot showing ‘S’ pangenome-based markers associated with TSWV Capture phenotypes. Capture represents the most susceptible phenotype observed throughout all

replicates and seasons. A Bonferroni-corrected *P* value of 0.05 was used as a significant threshold ($P = 4.4x10^{-7}$), represented by a horizontal dashed black line. The red dash line represents a Bonferroni-corrected *P* value of 0.001 ($P = 8.85x10^{-9}$). **b**, Local association plot of Chr.01 showing a clear signal. **c**, 'S' pangenome marker r^2 values of Chr.01 showing increased linkage disequilibrium throughout the pericentromeric region. **d**, Subset of 'S' RILs with pangenome marker calls for Chr.01 with founder origin coloration. Orange represents markers from the TSWV resistant founder NC94022 (NC).

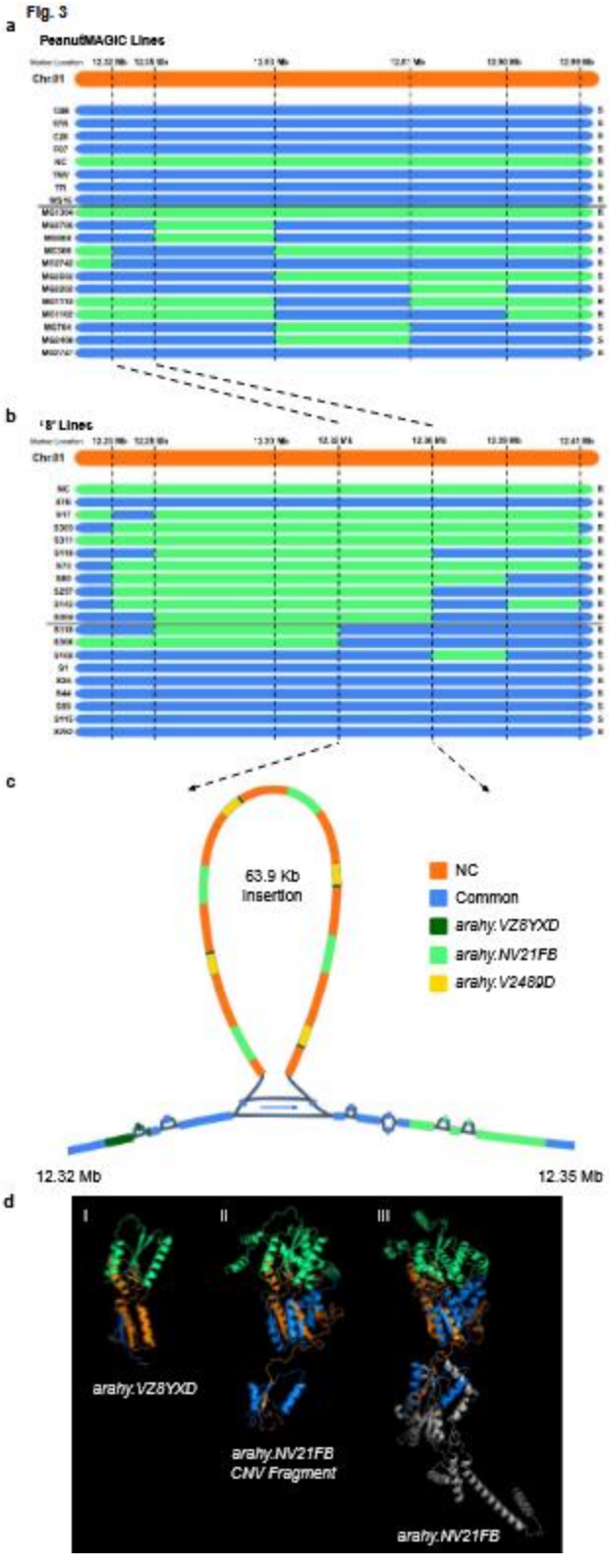


Fig. 3
a
Peanut8MAGIC Lines
Chr01
b
'8' Lines
Chr01
c
63.9 Kb
Insertion
NC
Common
arahy.VZ8YXD
arahy.NV21FB
arahy.V2489D
12.32 Mb
12.35 Mb
d
I
II
III
arahy.VZ8YXD
arahy.NV21FB
CNV Fragment
arahy.NV21FB

Fig. 4.3**: Copy Number Variant (CNV) within TSWV associated region**. **a**, Haplotypes of PeanutMAGIC RILs with NC founder specific markers. Green bars represent regions that stem from the resistant founder 'NC94022' (NC). Blue bars represent regions that come from a susceptible founder. MG306 possesses most of the region from NC while being susceptible. **b**, Haplotypes of 'S' RILs. Green bars represent regions that stem from the resistant parent 'NC94022' (NC). Blue bars represent regions that come from the susceptible parent 'SunOleic 97R' (97R). Lines with the region from 12.32 -12.36 that originate from NC express resistance where lines that have the region from 97R express susceptibility. **c**, Pangenomic visualization of the region isolated showed a 63.9 Kb CNV from the resistant found NC. Within the CNV there is a kinetochore protein from Chr.20 and sections of protein from neighboring glutamate receptors. **d**, Alphafold2 predictions of (I) *arahy.VZ8YXD*, (II) the glutamate receptor fragment within the CNV, and (III) *arahy.NV21FB*. Green, orange, blue regions highlight residues 1-100, 101-200, and 201-300, respectively.

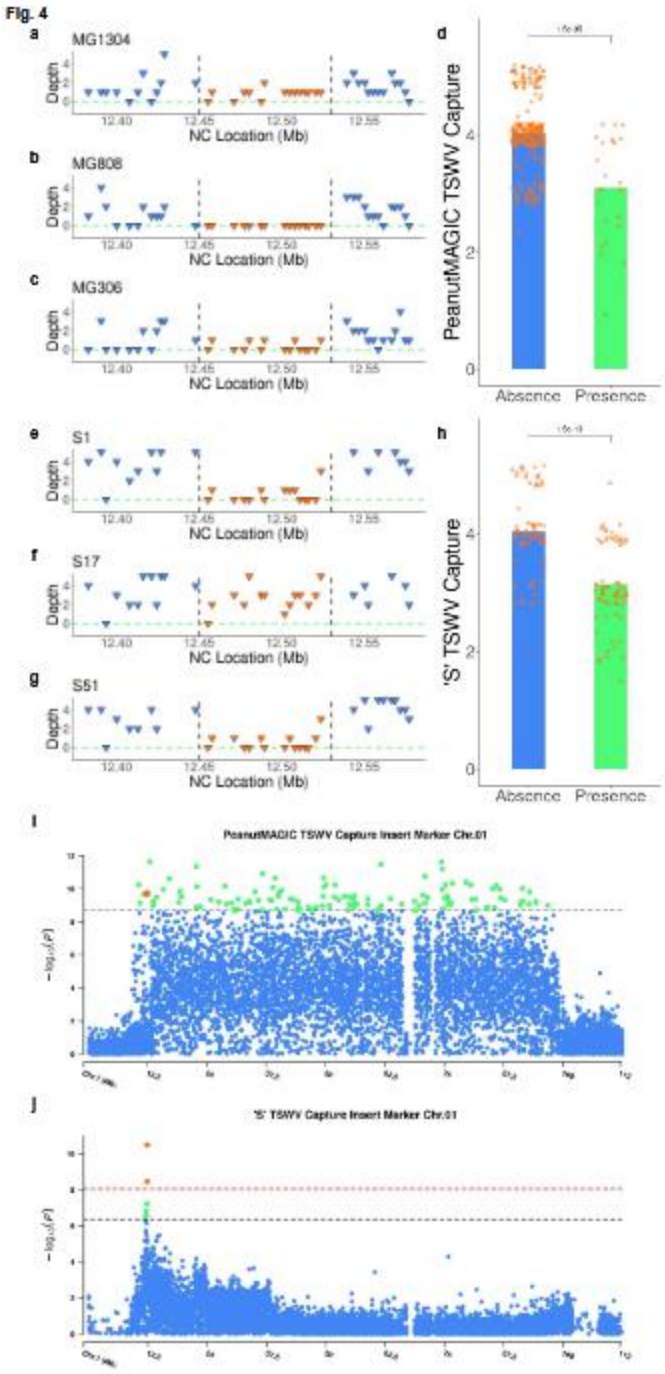


Fig. 4.4: **Personalized pangenome Copy Number Variant (CNV) detection**. **a-c**, Sequence depth of PeanutMAGIC RILs showing depth over the CNV region on Chr.01 for MG1304 (**a**), MG808 (**b**), and MG306 (**c**). The black vertical dashed lines represent the start and end of the CNV. The green horizontal dashed line highlights zero depth. **d**, MAGIC Core CNV

variance in comparison to Capture phenotypes. Statistical significance was calculated using a Student's *t*-test. The orange jitter plot represents individual phenotypes. **e-f**, Sequence depth of 'S' RILs showing depth over the CNV region on Chr.01 for S1 (**e**), S17 (**f**), and S51 (**g**). The black vertical dashed lines represent the start and end of the CNV. The green horizontal dashed line highlights zero depth. **h**, 'S' CNV variance in comparison to Capture phenotypes. Statistical significance was calculated using a Student's *t*-test. The orange jitter plot represents individual phenotypes. **i**, GWAS of Capture phenotypes and MAGIC Core CNV marker. The orange highlight represents the CNV marker. **j**, GWAS of Capture phenotypes and 'S' population CNV marker. The highest orange highlight represents the CNV marker.

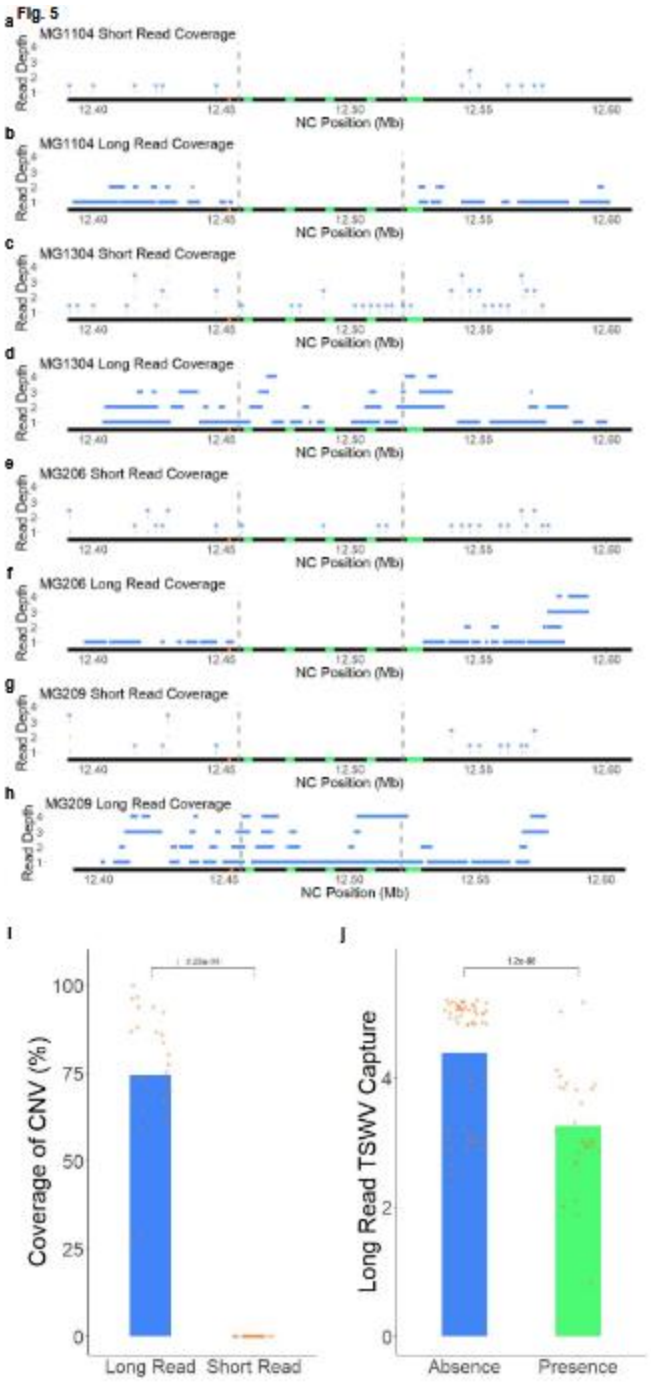


Fig. 4.5: **Long read low coverage Copy Number Variant (CNV) detection**. **a-b**, Short and long read sequence coverage over the CNV region for MG1104 (**a** & **b**), MG1304 (**c** & **d**), MG206 (**e** & **f**), and MG209 (**g** & **h**), respectively. The black horizontal bar represents the CNV region on Chr.01. The orange box represents *arahy.VZ8YXD*. The green horizontal

bars represent CNV *arahy.NV21FB* fragments and *arahy.NV21FB.* The blue horizontal bars represent sequence reads. The blue asterisks highlight the short read locations and depth. The gray vertical bars represent the start and end of the CNV. **i**, Comparison of short and long read coverage over the CNV for lines possessing the CNV in the long read data set. Statistical significance was calculated using a Student's *t*-test. The orange dots represent individual genotypes. **j**, Long read low coverage CNV variance in comparison to Capture phenotypes. Statistical significance was calculated using a Student's *t*-test. The orange jitter plot represents individual phenotypes.

Supplementary Table 4.1: The 45 MAGIC Core lines with 'NC94022' unique pericentromere (Peri.) compared to associated PeanutMAGIC pangenome-based markers.

Chr	Position	Lines with marker and NC Peri.	Lines with NC Peri. and missing marker	NC Marker and Resistant	NC Marker and Susceptible	Non-NC Marker
Chr.01	12,901,454	9	36	8	0	1
Chr.01	74,400,207	34	11	16	8	10
Chr.01	61,591,965	37	8	19	10	8
Chr.01	22,719,922	29	16	16	6	7
Chr.01	74,543,714	34	11	17	10	7
Chr.01	36,691,448	29	16	16	6	7
Chr.01	70,920,324	30	15	18	8	4
Chr.01	49,566,481	23	22	13	6	4
Chr.01	39,352,913	31	14	19	9	3
Chr.01	17,022,791	28	17	15	8	5
Chr.01	79,528,235	21	24	13	3	5
Chr.01	76,658,335	13	32	8	2	3

Chr.01	50,073,392	33	12	16	9	8
Chr.01	10,631,361	30	15	13	4	13
Chr.01	75,215,885	25	20	12	7	6

Supplementary Table 4.2: *arahy.V2489D* protein sequence blastp hits.

Description	Scientific Name	Query Cover	E Value	Per. Ident
kinetochore protein NDC80 homolog	*Arachis hypogaea*	100%	0	86.09
hypothetical protein AAHE18_20G006300	*Arachis hypogaea*	100%	0	85.94
probable kinetochore protein ndc80	*Arachis ipaensis*	100%	0	85.78
kinetochore protein NDC80 homolog isoform X1	*Arachis ipaensis*	100%	0	72.8
kinetochore protein NDC80 homolog	*Arachis hypogaea*	100%	0	72.64
hypothetical protein S83_069446	*Arachis hypogaea*	100%	0	72.49
kinetochore protein NDC80 homolog	*Arachis hypogaea*	100%	0	72.57
kinetochore protein NDC80 homolog isoform X1	*Arachis ipaensis*	100%	0	72.57
hypothetical protein S245_069711	*Arachis hypogaea*	100%	0	72.42
kinetochore protein NDC80 homolog	*Arachis hypogaea*	100%	0	72.02
kinetochore protein NDC80 homolog	*Arachis duranensis*	100%	0	71.67

kinetochore protein NDC80 homolog	*Arachis stenosperma*	100%	0	71.65
kinetochore protein NDC80 homolog	*Arachis stenosperma*	100%	0	71.56
hypothetical protein Ahy_B10g101356	*Arachis hypogaea*	100%	0	71.03
hypothetical protein Ahy_A10g050724	*Arachis hypogaea*	100%	0	71.16
hypothetical protein	*Stylosanthes scabra*	100%	0	69.84
hypothetical protein	*Stylosanthes scabra*	100%	0	70.32
hypothetical protein	*Stylosanthes scabra*	98%	0	68.4
hypothetical protein	*Stylosanthes scabra*	98%	0	67.92
kinetochore protein NDC80 homolog	*Arachis duranensis*	86%	0	73.15
kinetochore protein NDC80 homolog isoform X2	*Arachis ipaensis*	78%	0	83.37
hypothetical protein S245_069712	*Arachis hypogaea*	73%	0	83.05
kinetochore protein NDC80 homolog isoform X1	*Arachis ipaensis*	72%	0	83.44
kinetochore protein NDC80 homolog isoform X2	*Arachis hypogaea*	72%	0	83.44
kinetochore protein NDC80 homolog isoform X2	*Arachis ipaensis*	73%	0	82.83

Supplementary Table 4.3: CNV putative kinetochore protein sequence blastp hits.

Description	Scientific Name	Query Cover	E Value	Per. Ident
hypothetical protein S245_069712	*Arachis hypogaea*	78%	1.00E-157	90.23
kinetochore protein NDC80 homolog isoform X2	*Arachis hypogaea*	77%	6.00E-157	91.16
kinetochore protein NDC80 homolog isoform X1	*Arachis ipaensis*	77%	7.00E-157	91.16
kinetochore protein NDC80 homolog isoform X2	*Arachis ipaensis*	78%	7.00E-157	89.84
hypothetical protein AAHE18_20G006300	*Arachis hypogaea*	77%	5.00E-156	93.98
kinetochore protein NDC80 homolog	*Arachis hypogaea*	77%	5.00E-156	93.98
probable kinetochore protein ndc80	*Arachis ipaensis*	77%	5.00E-156	93.98
kinetochore protein NDC80 homolog	*Arachis duranensis*	77%	1.00E-154	90.36
kinetochore protein NDC80 homolog isoform X2	*Arachis ipaensis*	78%	3.00E-154	88.98
kinetochore protein NDC80 homolog	*Arachis stenosperma*	77%	5.00E-154	90.76
kinetochore protein NDC80 homolog isoform X2	*Arachis ipaensis*	77%	8.00E-154	90.36
kinetochore protein NDC80 homolog	*Arachis hypogaea*	77%	9.00E-154	90.36
kinetochore protein NDC80 homolog	*Arachis hypogaea*	77%	1.00E-153	90.36

kinetochore protein NDC80 homolog isoform X1	*Arachis ipaensis*	77%	1.00E-153	90.36
hypothetical protein S83_069446	*Arachis hypogaea*	77%	3.00E-153	90.36
hypothetical protein S245_069711	*Arachis hypogaea*	77%	5.00E-153	90.36
kinetochore protein NDC80 homolog isoform X1	*Arachis ipaensis*	77%	6.00E-153	90.36
kinetochore protein NDC80 homolog	*Arachis duranensis*	77%	3.00E-149	89.56
kinetochore protein NDC80 homolog	*Arachis hypogaea*	77%	8.00E-149	89.16
kinetochore protein NDC80 homolog	*Arachis stenosperma*	77%	1.00E-147	88.76
hypothetical protein Ahy_A10g050724	*Arachis hypogaea*	77%	2.00E-144	83.77
hypothetical protein Ahy_B10g101356	*Arachis hypogaea*	77%	6.00E-143	86.06
hypothetical protein	*Stylosanthes scabra*	77%	5.00E-141	85.94
hypothetical protein	*Stylosanthes scabra*	77%	2.00E-140	85.94
hypothetical protein	*Stylosanthes scabra*	77%	1.00E-139	85.94

Supplementary Table 4.4: *arahy.NV21FB* protein sequence blastp hits.

Description	Scientific Name	Query cover	E Value	Per. Ident
hypothetical protein HN51_000890	*Arachis hypogaea*	100%	0	95.92
glutamate receptor 3.6-like	*Arachis duranensis*	100%	0	95.25
glutamate receptor 3.6-like	*Arachis hypogaea*	89%	0	95.44
glutamate receptor 3.6-like isoform X1	*Arachis stenosperma*	100%	0	85.73
hypothetical protein Ahy_B10g100359	*Arachis hypogaea*	100%	0	84.57
Glutamate receptor 3	*Arachis hypogaea*	100%	0	84.59
hypothetical protein AAHE18_20G006200	*Arachis hypogaea*	100%	0	84.59
glutamate receptor 3.6-like	*Arachis stenosperma*	100%	0	86.01
hypothetical protein S245_068869	*Arachis hypogaea*	100%	0	82.28
glutamate receptor 3.6-like	*Arachis duranensis*	89%	0	88.56
hypothetical protein VNO77_02318	*Canavalia gladiata*	100%	0	75.72
Glutamate receptor 3.6	*Spatholobus suberectus*	100%	0	75.61
glutamate receptor 3.6-like	*Gastrolobium bilobum*	100%	0	76.38

hypothetical protein VNO77_02314	*Canavalia gladiata*	100%	0	74.72
glutamate receptor 3.6	*Glycine max*	100%	0	75.61
hypothetical protein JHK86_025776	*Glycine max*	100%	0	75.39
hypothetical protein AAZX31_09G180100	*Glycine max*	100%	0	75.39
hypothetical protein JHK82_025646	*Glycine max*	100%	0	74.94
glutamate receptor 3.6	*Glycine max*	100%	0	74.94
hypothetical protein JHK85_026266	*Glycine max*	98%	0	75.56
Glutamate receptor 3.3	*Glycine max*	100%	0	74.72
glutamate receptor 3.6-like	*Abrus precatorius*	100%	0	73.62
glutamate receptor 3.6	*Cajanus cajan*	100%	0	74.17
Glutamate receptor 3.3	*Glycine max*	100%	0	74.17
glutamate receptor 3.6	*Cajanus cajan*	100%	0	73.32

Supplementary Table 4.5: *arahy.VZ8YXD* protein sequence blastp hits.

Description	Scientific Name	Query Cover	E Value	Per. Ident
hypothetical protein S83_000879	*Arachis hypogaea*	100%	2.00E-166	100
glutamate receptor 3.3-like	*Arachis hypogaea*	89%	2.00E-144	100
hypothetical protein Ahy_A01g003846	*Arachis hypogaea*	87%	4.00E-131	95.45
glutamate receptor 3.6-like	*Arachis stenosperma*	85%	3.00E-124	97.42
glutamate receptor 3.6-like	*Arachis duranensis*	85%	2.00E-109	88.14
hypothetical protein HN51_000890	*Arachis hypogaea*	85%	3.00E-109	88.14
glutamate receptor 3.6-like isoform X1	*Arachis stenosperma*	85%	6.00E-108	87.11
Glutamate receptor 3	*Arachis hypogaea*	85%	1.00E-107	87.11
hypothetical protein AAHE18_20G006200	*Arachis hypogaea*	85%	1.00E-107	87.11
Glutamate receptor 3.6	*Spatholobus suberectus*	85%	1.00E-105	86.08
glutamate receptor 3.6-like	*Gastrolobium bilobum*	85%	1.00E-104	87.11
glutamate receptor 3.6	*Cajanus cajan*	85%	5.00E-104	84.02
hypothetical protein S245_068869	*Arachis hypogaea*	85%	5.00E-104	87.11

glutamate receptor 3.6	*Glycine max*	85%	7.00E-104	84.02
hypothetical protein Ahy_B10g100359	*Arachis hypogaea*	85%	8.00E-104	87.11
glutamate receptor 3.6	*Cajanus cajan*	85%	2.00E-103	84.02
hypothetical protein JHK86_025776	*Glycine max*	85%	3.00E-103	83.51
hypothetical protein AAZX31_09G180100	*Glycine max*	85%	4.00E-103	83.51
glutamate receptor 3.3-like	*Gastrolobium bilobum*	85%	5.00E-103	82.99
glutamate receptor 3.6-like	*Abrus precatorius*	85%	9.00E-103	84.02
hypothetical protein VNO77_02318	*Canavalia gladiata*	85%	2.00E-101	85.05
glutamate receptor 3.6	*Vigna angularis*	85%	1.00E-100	81.96
hypothetical protein V8G54_000907	*Vigna mungo*	85%	2.00E-100	81.96
hypothetical protein VNO77_02314	*Canavalia gladiata*	85%	3.00E-100	80.41
glutamate receptor 3.6-like isoform X2	*Vigna unguiculata*	85%	4.00E-100	81.96

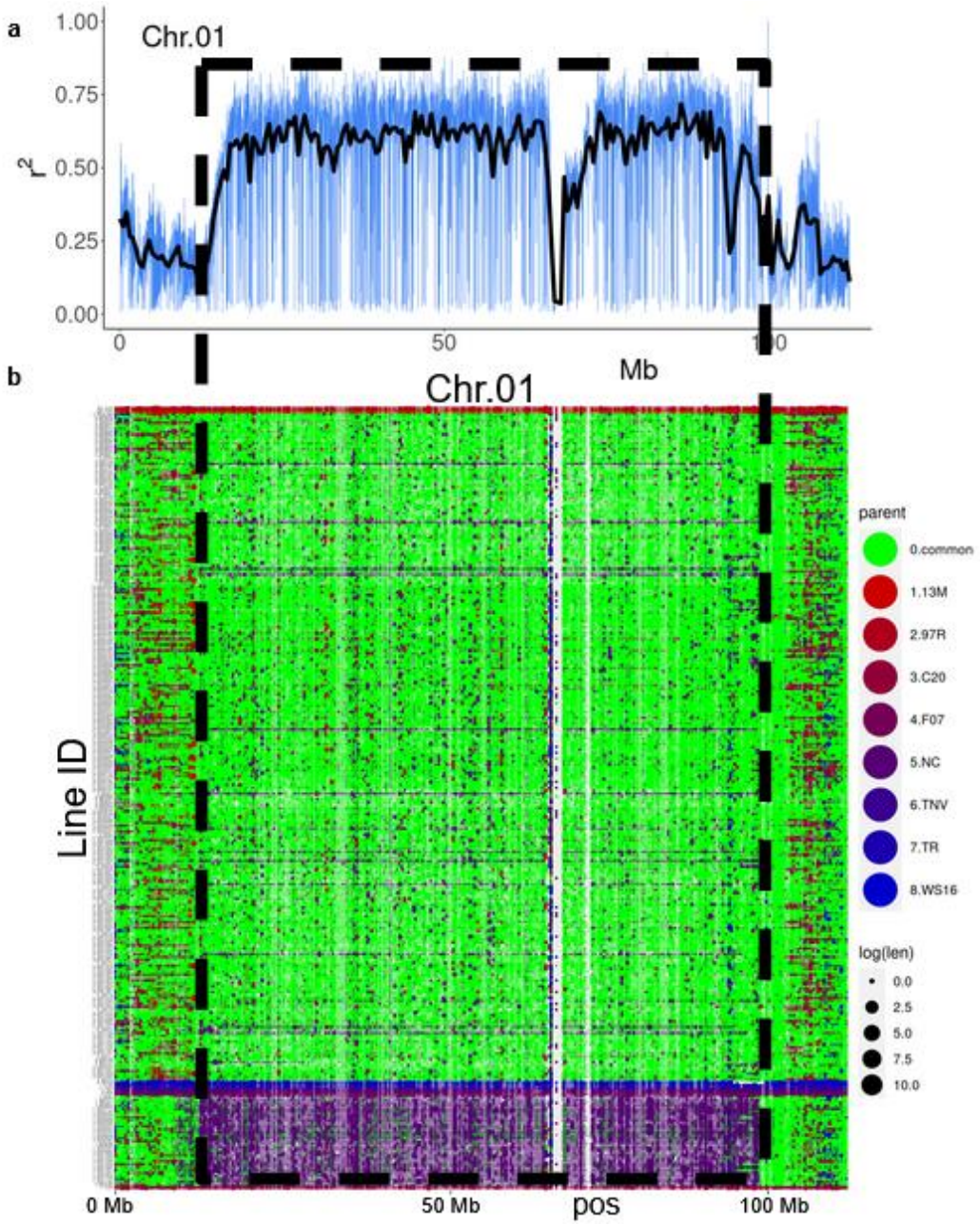
Supplementary Table 4.6: CNV putative glutamate receptor protein sequence blastp hits.

Description	Scientific Name	Query Cover	E Value	Per. Ident
hypothetical protein HN51_000890	*Arachis hypogaea*	100%	0	100
glutamate receptor 3.6-like	*Arachis duranensis*	100%	0	99.09
glutamate receptor 3.6-like	*Arachis stenosperma*	100%	0	93.51
glutamate receptor 3.6-like	*Arachis hypogaea*	83%	0	100
glutamate receptor 3.6-like isoform X1	*Arachis stenosperma*	100%	0	83.84
glutamate receptor 3.6-like	*Arachis duranensis*	83%	0	99.13
hypothetical protein Ahy_B10g100359	*Arachis hypogaea*	100%	0	82.26
hypothetical protein S245_068869	*Arachis hypogaea*	100%	0	82.26
Glutamate receptor 3	*Arachis hypogaea*	100%	0	82.26
hypothetical protein AAHE18_20G006200	*Arachis hypogaea*	100%	0	82.26
Glutamate receptor 3.6	*Spatholobus suberectus*	100%	0	77.72
hypothetical protein VNO77_02318	*Canavalia gladiata*	100%	0	77.72
glutamate receptor 3.6	*Glycine max*	100%	0	77.9
hypothetical protein JHK86_025776	*Glycine max*	100%	0	77.54

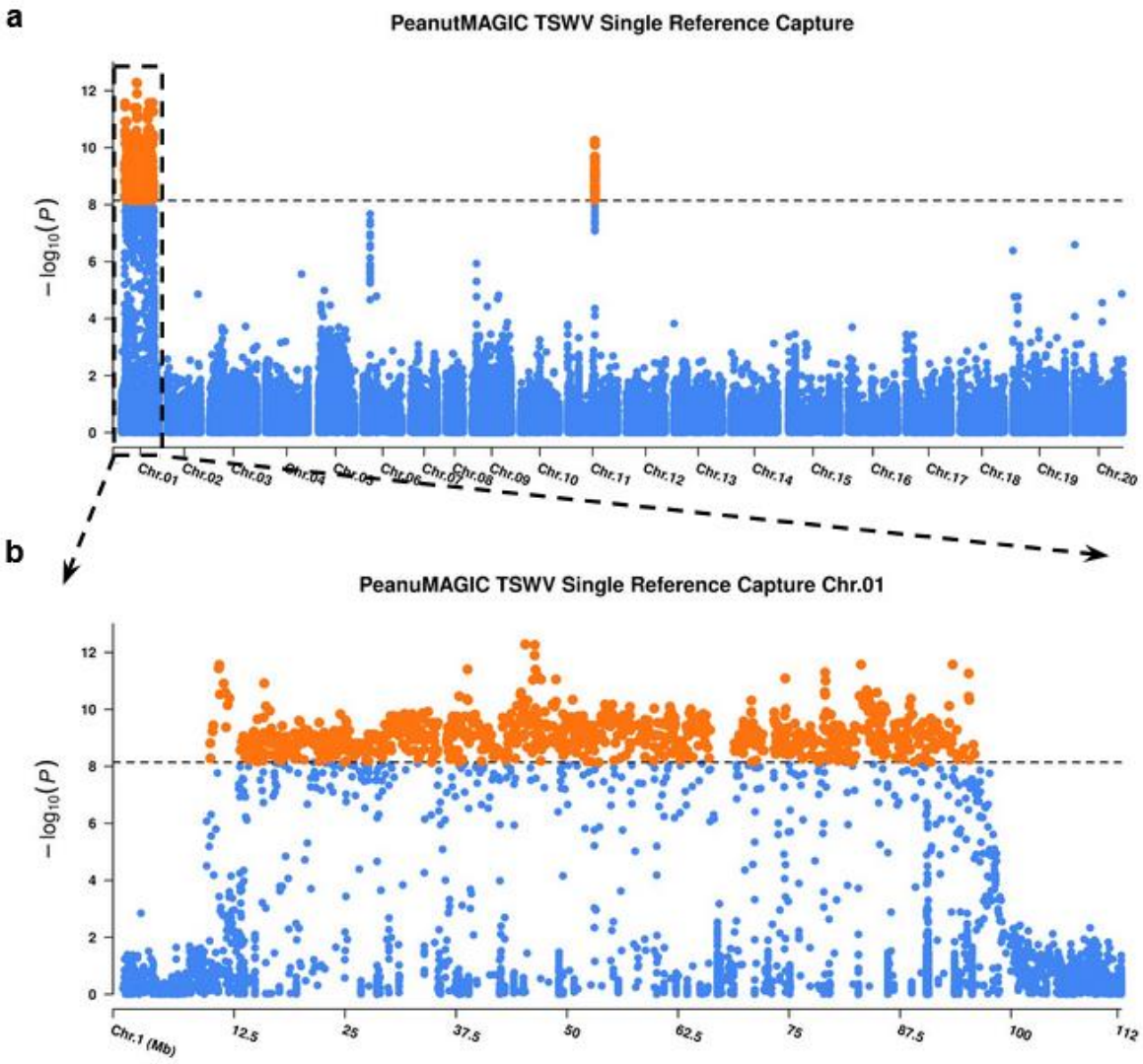
hypothetical protein AAZX31_09G180100	*Glycine max*	100%	0	77.54
hypothetical protein VNO77_02314	*Canavalia gladiata*	100%	0	75.91
glutamate receptor 3.6-like	*Gastrolobium bilobum*	100%	0	78.08
hypothetical protein JHK85_026266	*Glycine max*	98%	0	77.86
glutamate receptor 3.3-like	*Gastrolobium bilobum*	100%	0	76.23
Glutamate receptor 3.3	*Glycine max*	100%	0	76.45
hypothetical protein JHK82_025646	*Glycine max*	100%	0	75.59
glutamate receptor 3.6	*Glycine max*	100%	0	75.59
Glutamate receptor 3.3 isoform B	*Glycine soja*	100%	0	75.59
glutamate receptor 3.6	*Cajanus cajan*	100%	0	75.41
glutamate receptor 3.6-like	*Abrus precatorius*	100%	0	74.46

Supplementary Table 4.7: MAGIC Core lines TSWV phenotypes over three seasons. 1 represents no apparent symptomology and 5 represents severe stunting, spotting, and chlorosis.

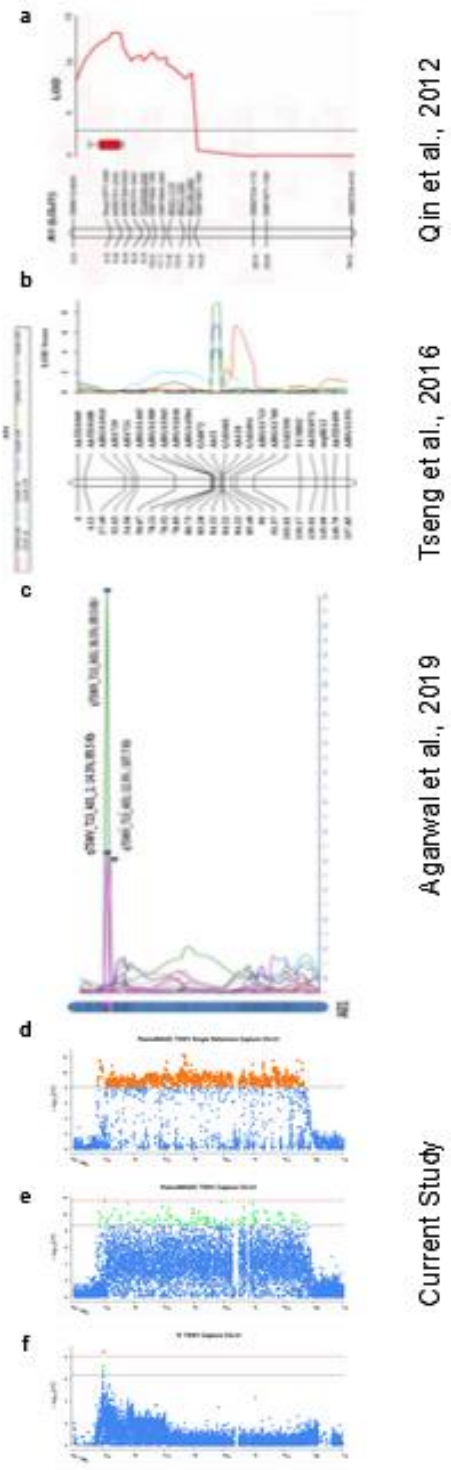
Genotype	MG1106	MG12	MG1207	MG1301	MG1302	MG1306	MG2009	MG206	MG2402	MG2406	MG2409	MG2502	MG2704
Insert	Presence	Presence	Presence	Presence	Presence	Presence	Presence	Presence	Presence	Presence	Presence	Presence	Presence
TSWV_21	1	1	4	1	1	4	1	1	2	2	2	2	4
TSWV_Rep1_22	3	4	2	2	3	2	2	2	3	3	1	2	2
TSWV_Rep2_22	1	2	4	3	3	4	1	3	1	1	2	3	2
TSWV_Rep3_22	2	2	1	2	2	2	3	1	3	3	1	2	2
TSWV_AVG22	2	2.67	2.33	2.33	2.67	2.67	2	2	2.33	2.33	1.33	2.33	2
TSWV_Rep1_23	2	3	1.5	2.5	3.5	4	2	4	3.5	4	3	4	1
TSWV_Rep2_23	1	2	2	1	1	2	2	2	1	2	2	2	3
TSWV_Rep3_23	2	1	1	1	1	1	1	2	1	3	2	2	2
TSWV_Rep4_23	1	1	2	2	2	2	2	2	1	1	1	1	3
TSWV_AVG_23	1.5	1.75	1.625	1.625	1.875	2.25	1.75	2.5	1.625	2.5	2	2.25	2.25
TSWV_YR_AVG	1.5	1.81	2.65	1.65	1.85	2.97	1.58	1.83	1.99	2.28	1.78	2.19	2.75
TSWV_Rep_AVG	1.625	2	2.1875	1.8125	2.0625	2.625	1.75	2.125	1.9375	2.375	1.75	2.25	2.375
TSWV_Capture	3	4	4	3	3.5	4	3	4	3.5	4	3	4	4



Supplementary Fig. 4.1: Characteristics of Chr.01 recombination of MAGIC Core. **a**, r^2 values across Chr.01 calculated from a sliding window of 300 markers, depicting the pericentromeric region with elevated LD. The black line represents the average r^2 value in 0.5 Mb. **b**, MAGIC Core pangenome marker calls for Chr.01 with founder origin coloration. The black dashed box highlights the pericentromeric region with increased LD stemming from NC.

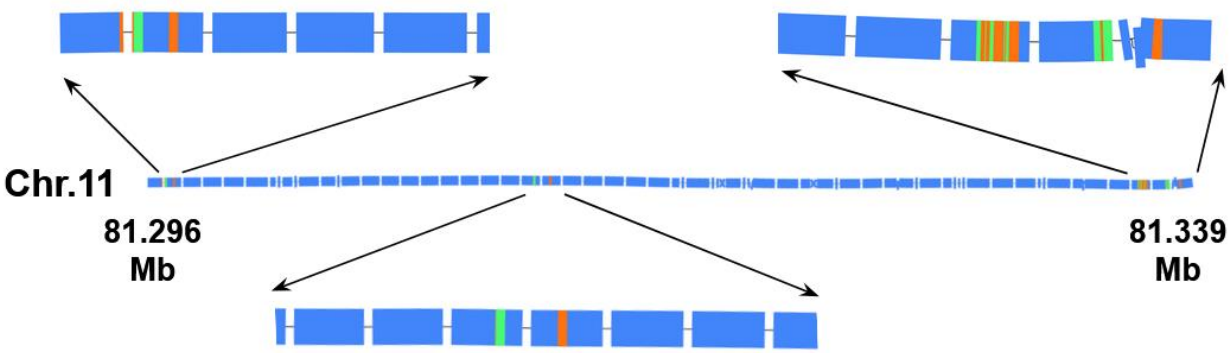


Supplementary Fig. 4.2: Single reference-based marker association with TSWV Capture phenotypes. **a**, Whole genome Manhattan plot of association signal. The significant threshold is a Bonferroni-corrected *P* value of 0.001 ($P = 7.24 \times 10^{-9}$). Significant associations were identified on Chr.01 and Chr.11. **b**, Zoom-in of Chr.01 association signal.

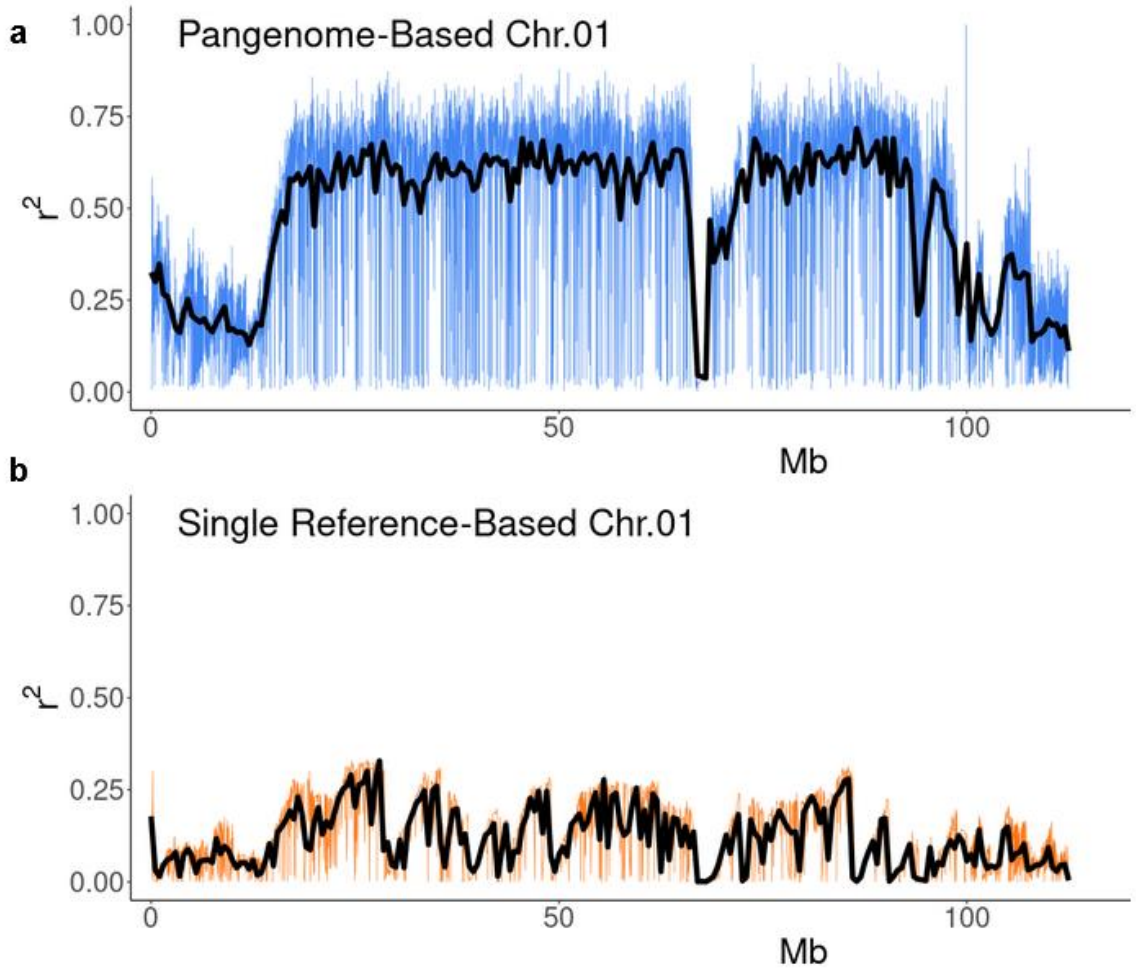


Supplementary Fig. 4.3: TSWV QTL mapping from Qin et al., 2012 (a) Tseng et al., 2016 (b) and Agarwal et al., 2019 (c) compared to MAGIC Core single refence-based marker association

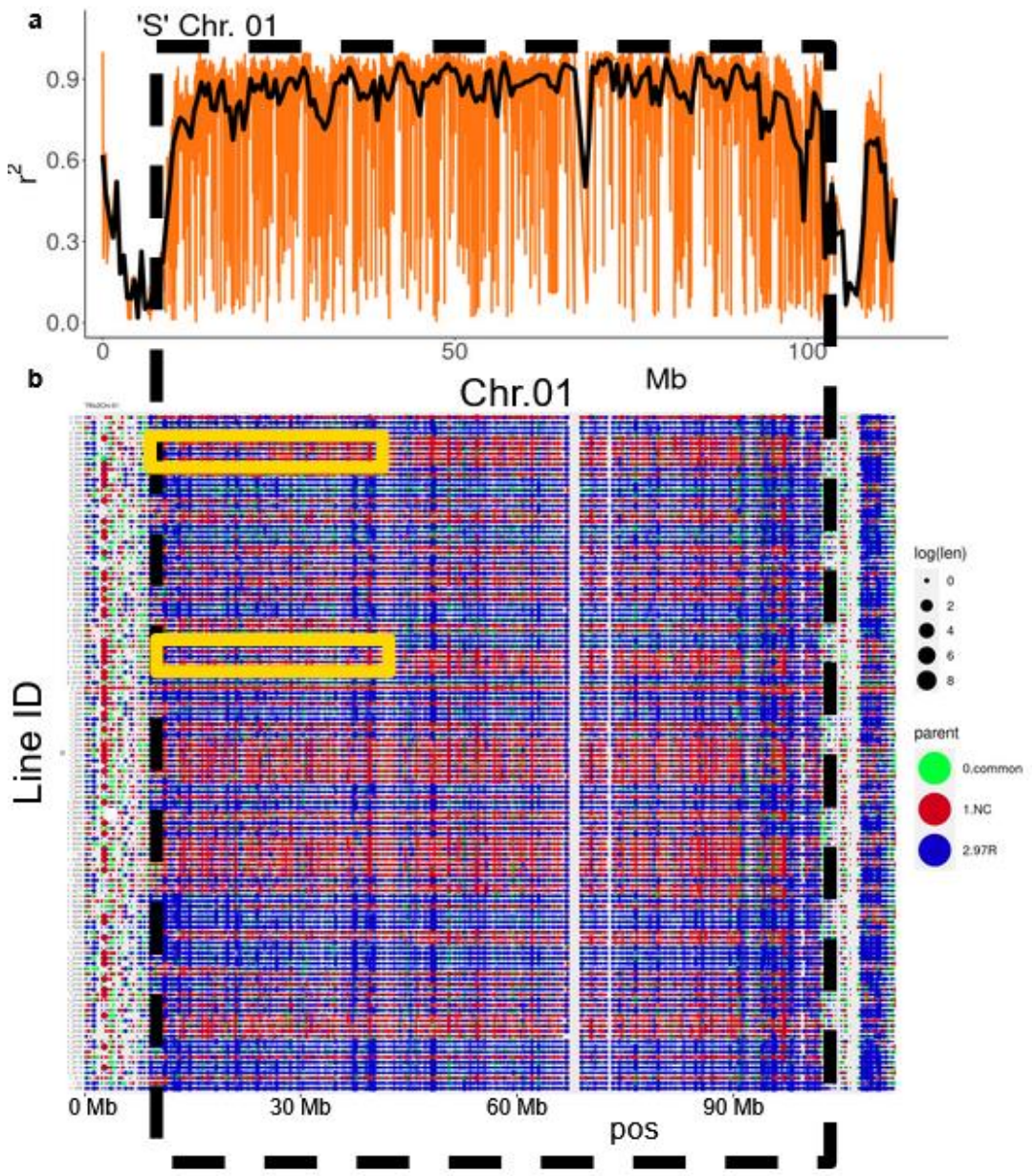
(d), MAGIC Core pangenome-based marker association (e), and ‘S’ population pangenome-based marker association (f).



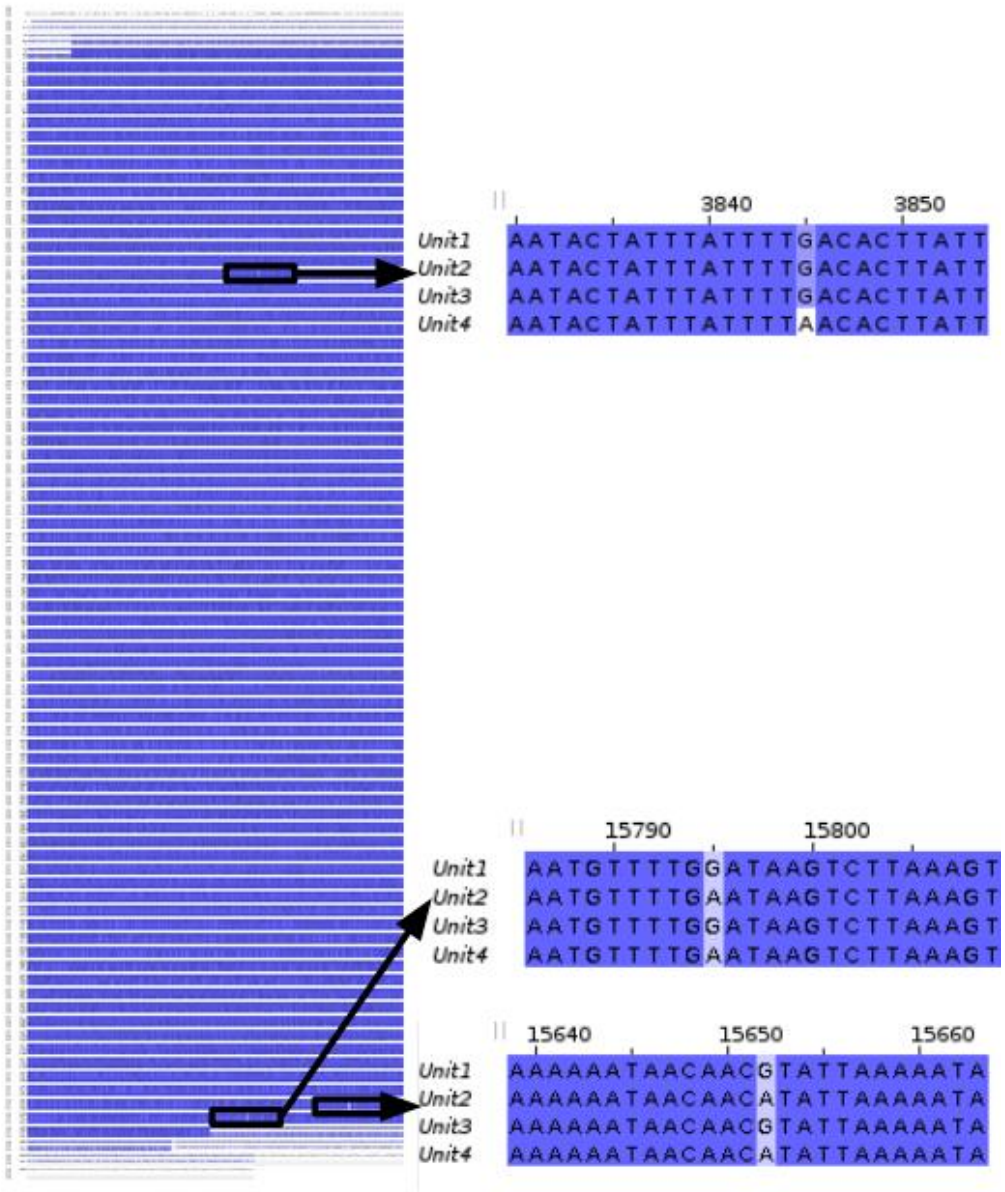
Supplementary Fig. 4.4: PeanutMAGIC pangenome visualization of the single-reference TSWV Capture associated region on Chr.11, spanning 43 Kb (81,296,099 – 81,339,263). There are 25 single-reference-based markers within the region and are highlighted by orange and green bars that are in 131 bp blocks to represent the size of the read. Note some blocks are overlapped with other blocks. These regions do not align with variants in the region of the pangenome.



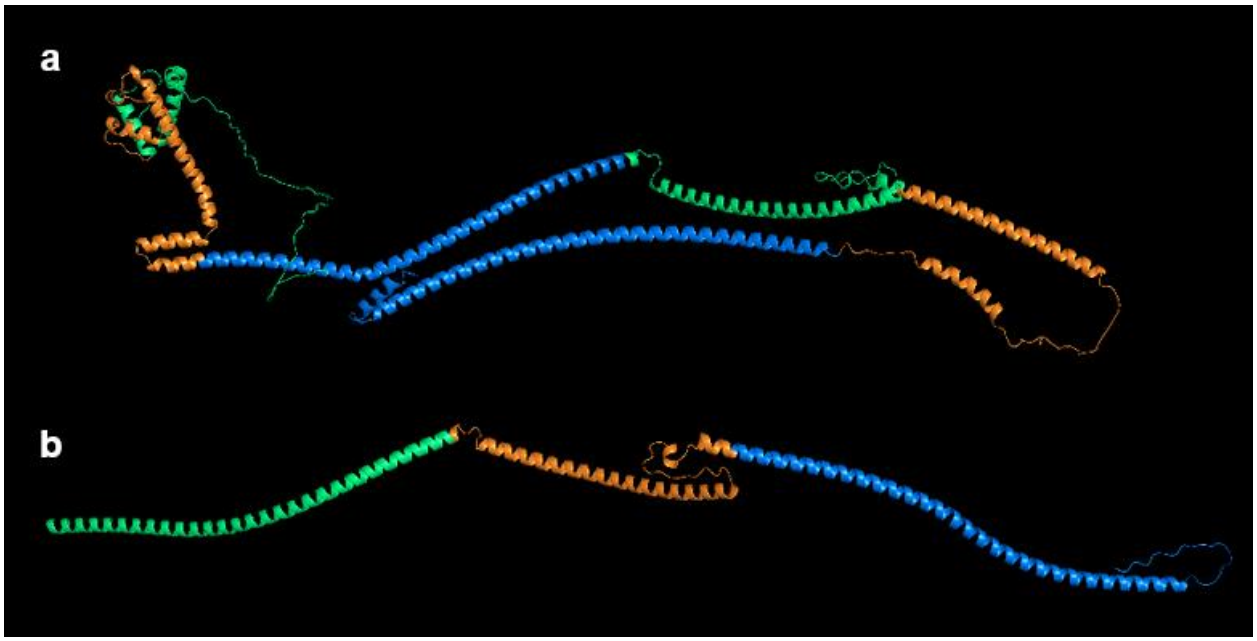
Supplementary Fig. 4.5: Comparison of Chr.01 recombination with PeanutMAGIC pangenome based markers and single reference-based markers. **a**, r^2 values across Chr.01 calculated from a sliding window of 300 pangenome-based markers. The black line represents the average r^2 value in 0.5 Mb. **b**, Figure 4b from Thompson et al. (2024) depicting the r^2 values across Chr.01 calculated from a sliding window of 100 single reference markers and averaged in 0.5 Mb bins. The increased number of markers in the sliding window for PeanutMAGIC pangenome markers is due to the increased marker density, allowing for equal physical window sizes between single reference and pangenome-based markers.



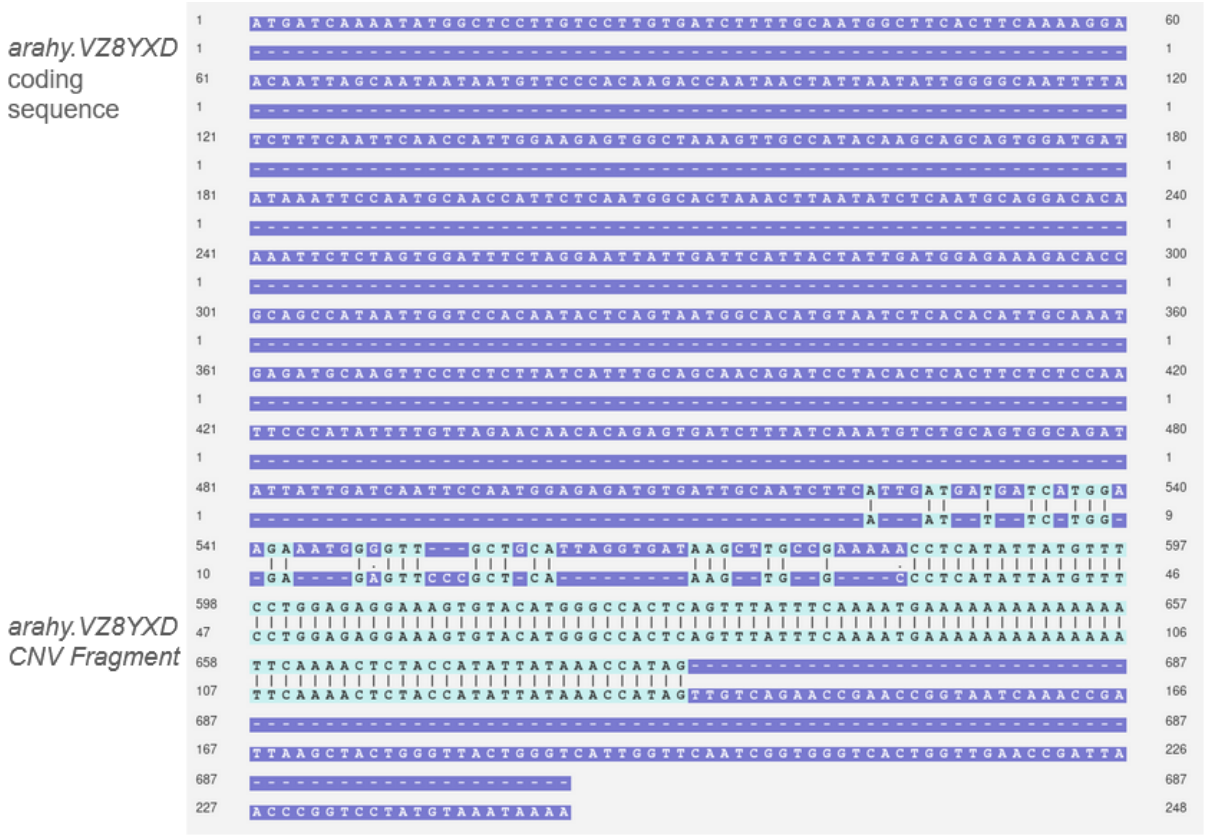
Supplementary Fig. 4.6: Characteristics of Chr.01 recombination of 'S' population. **a**, r^2 values across Chr.01 calculated from a sliding window of 100 markers, depicting the pericentromeric region with elevated LD. The black line represents the average r^2 value in 0.5 Mb. **b**, 'S' pangenome marker calls for Chr.01 with founder origin coloration. The black dashed box highlights the pericentromeric region with increased LD stemming from NC. The yellow box highlights large fractions of NC pericentromeric region recombination.



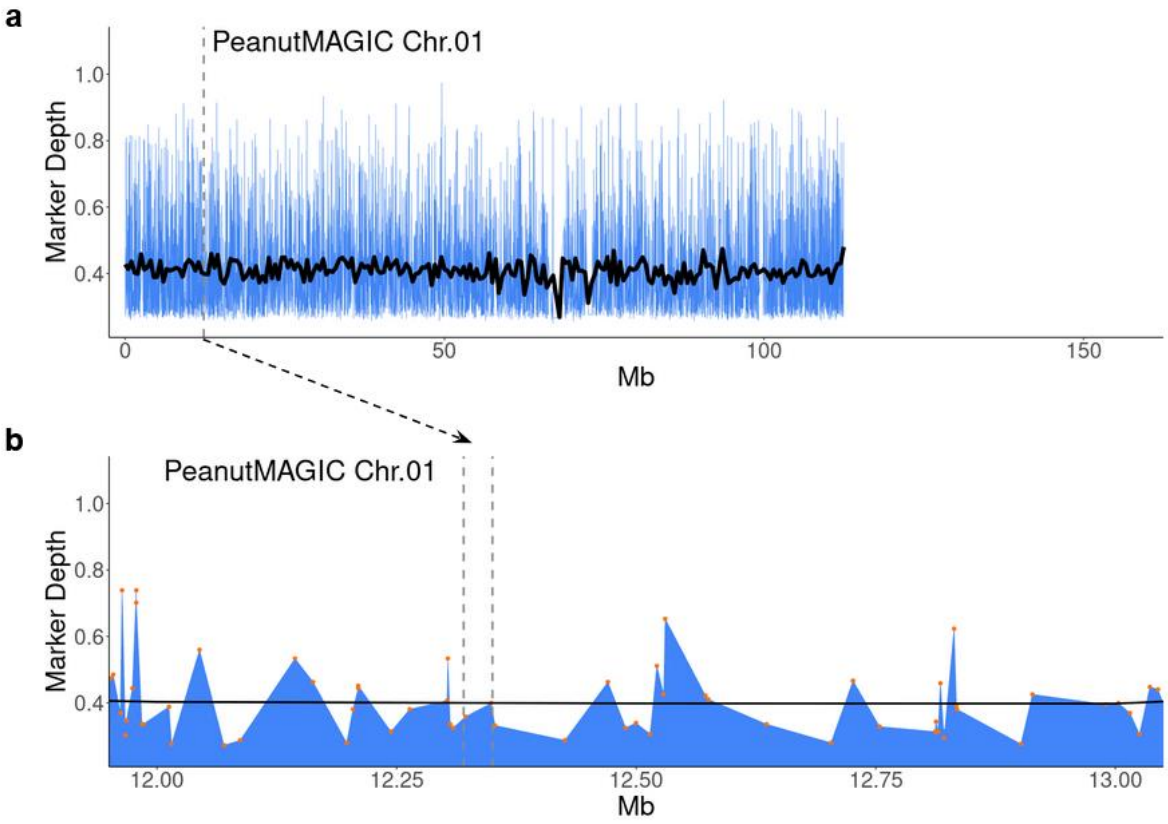
Supplementary Fig. 4.7: Sequence alignment of the four units of the associated CNV. There are three A/G SNP sites of difference.



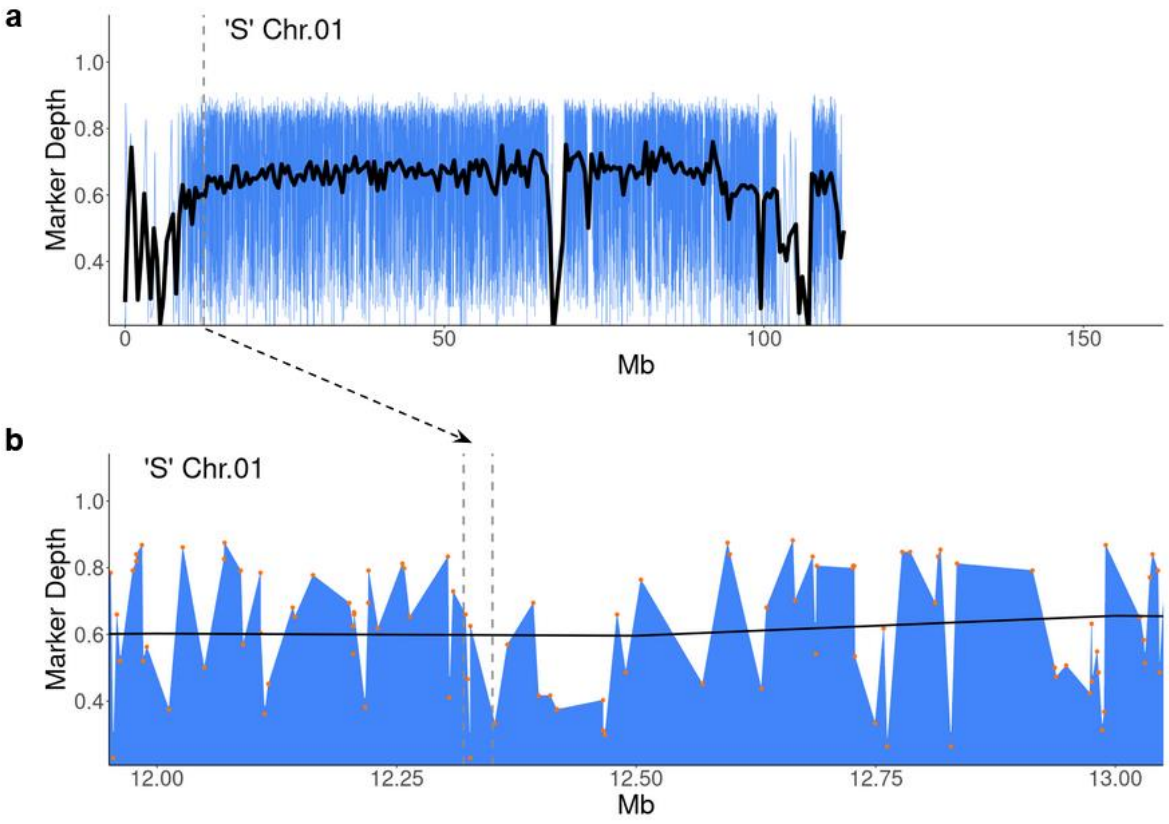
Supplementary Fig. 4.8: Alphafold2 protein prediction models of reference annotation kinetochore *arahy.V2489D* (a) and CNV kinetochore protein sequence.



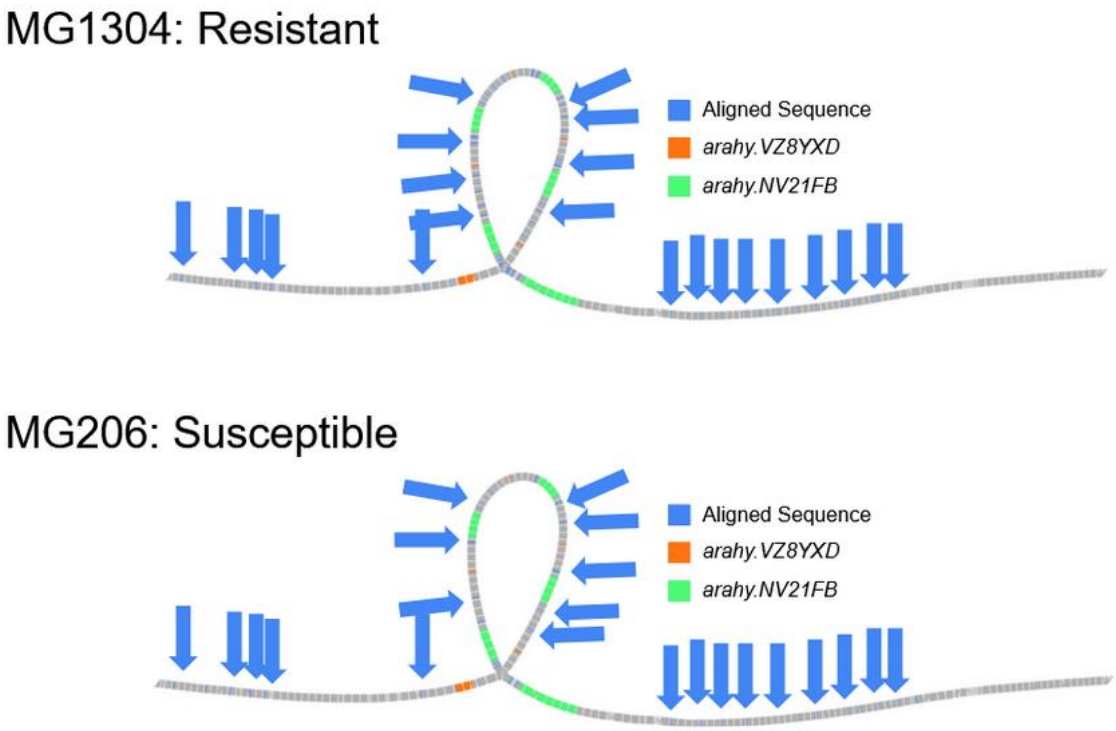
Supplementary Fig. 4.9: Sequence alignment of reference codon annotation *arahy.VZ8YXD* in line one and CNV fragment in line two.



Supplementary Fig. 4.10: Marker depth of Chr.01 in MAGIC Core. **a,** Marker depth throughout Chr.01 on a normalized scale from 0-1. The black line represents the average in 0.5 Mb. **b,** Zoom in of the associated region from Capture TSWV.



Supplementary Fig. 4.11: Marker depth of Chr.01 in 'S' population. **a,** Marker depth throughout Chr.01 on a normalized scale from 0-1. The black line represents the average in 0.5 Mb. **b,** Zoom in of the associated region from Capture TSWV.



Supplementary Fig. 4.12: CNV visualization with short read data aligned. The blue arrows highlight the full short read sequences that mapped to the region, making calling of the CNV difficult and error prone. Both MG1304 and MG206 were classified as possessing the CNV however MG206 was identified to not have the CNV using long read sequences.

CHAPTER 5

CONCLUDING REMARKS AND FUTURE PRESPECTIVES

This dissertation presents the development, genomic characterization, and application of the PeanutMAGIC and a population specific pangenome to enhance genetic diversity and empower high-resolution trait mapping in peanut. In chapter two we demonstrated that the PeanutMAGIC population is suitable for genomic studies through a subset termed the MAGIC Core. In chapter three we exemplified the difficulties to characterize the PeanutMAGIC population utilizing a single reference genome. The use of a population specific pangenome empowered genomic characterization and association studies to reveal a third FAD2 gene, resolving a mystery that has confused the peanut community. In chapter four we applied these approaches to identify a copy number variant that confers TSWV resistance that has eluded several previous efforts. These findings underscore the transformative potential of population-specific pangenomics in crop research. By indexing segregating variation within a defined breeding population, this approach mitigates reference bias, enhances marker resolution, and enables the discovery of structural variants with functional relevance. The PeanutMAGIC population, along with its associated genomic resources, stands as an asset for the peanut research community, offering new avenues for trait discovery, genetic improvement, and cultivar development. In conclusion, this work not only advances our understanding of peanut genomics but also provides a scalable framework for leveraging pangenomic tools in other complex crop systems. The integration of high-resolution mapping, structural variant analysis, and functional

genomics within a multiparent context sets a new standard for trait dissection and breeding innovation.

APPENDIX

INFORMATION OF LEAF SPOT AND ROOT KNOT NEMATODE ASSOCIATIONS USING POPULATION SPECIFIC PANGNEOME

Leaf Spot Phenotyping and Association

The MAGIC Core was planted in Tifton, GA for three years with three replicates per year. For each year the field was evaluated for early and late leaf spot (ELS, LLS, respectively) presence. It was found that in each year most lesions were late leaf spot, consistently throughout the ratings. The greatest percentage of early leaf spot identification was approximately 20 percent during the first rating of the 2023 season with many ratings not identifying ELS. The inconsistent low levels of ELS make observations difficult and were disregarded. The consistent high proportions of LLS offer a more direct study of the interactions between LLS and RILs. Each year had either three or four scores of each plot, allowing for the calculation of area under the disease progress curve (AUDPC). The AUDPC allows for a value to represent the progression of the disease on a particular plot throughout the season, opposed to a single time point. We performed a GWAS using the average of each plot across all replicates AUDPC (Fig. 6.1). We found no significant locations.

The results of the LLS GWAS may be the product of the dilution of resistance alleles in PeanutMAGIC RILs that may only combine in rare recombinants and the effects of these alleles are minimal. Additionally, leaf spot resistance in the PeanutMAGIC founders is not as evident as TSWV resistance and might be composed of multiple unlinked alleles, making their observations difficult in a multiparental population framework. The resistant founder is ‘GP-NC WS16’ which

has some level of resistance, however lesser than TSWV or root knot nematode. The sequencing of the full PeanutMAGIC population may offer insight into lines that can be selected to target genomic features that come from specific founders. In this case, genomic features unique to 'GP-NC WS16', particularly on Chr.05, could be identified and selected for in different RILs to be evaluated, offering a targeted population for leaf spot resistance based on previous studies. The MAGIC Core is a random subset of PeanutMAGIC to test if the population is suited for genomic studies and may not have specific allele frequencies needed to isolate a trait of interest. We hope to have these such sources available for the community to facilitate more collaborative research. Additionally, bulk analysis of equal parts of the most susceptible and most resistant RILs may offer insight into regions that have the most influence over leaf spot resistances. Together predictive selection and bulk analysis may identify alleles that can be useful for ELS and or LLS management.

Root Knot Nematode Phenotyping and Association

One of the founders of the PeanutMAGIC population is 'TifNV-High O/L' which possesses an *A. cardenasii* introgression that confers resistance to peanut root knot nematode (*Meloidogyne arenaria* race 1, RKN). We found the recombination of this introgression in the MAGIC Core was the source of the population structure (Fig. 6.2). Furthermore, we identified macro and micro fragments of the introgression in the MAGIC Core, that was previously undetected in previous biparental populations and in single reference MAGIC Core datasets, facilitating a more comprehensive understanding of the introgression and recombination (Fig. 6.3). To test if MAGIC Core recombinants could inherit RKN resistance we inoculated the 52 RILs in the introgression clade, 30 days after planting, with RKN (Holbrook, 1983). At 75 days after inoculation the lines were subject to egg extraction and counting. We found that lines

containing the introgression possessed resistance compared to susceptible controls, in parallel to 'TifNV-High O/L'. Furthermore, we found lines that possessed majority of the introgression, however expressed a susceptible phenotype. This finding showed that RKN resistance does segregate in the MAGIC Core and can disassociate from the introgression, in parallel to previous studies. We then performed inoculations on the whole MAGIC Core for association, following the previous RKN phenotyping approach. The egg counts were transformed using square root to fit the normal distribution assumption of GWAS. We identified a 400 Kb region on Chr.09 (Fig. 6.4). Although only one replicate of phenotypes has been performed, they were consistent between the 52 RILs and whole MAGIC Core. We hope to continue this study in the future to isolate RKN resistance gene(s) in peanut to facilitate improved cultivar breeding.

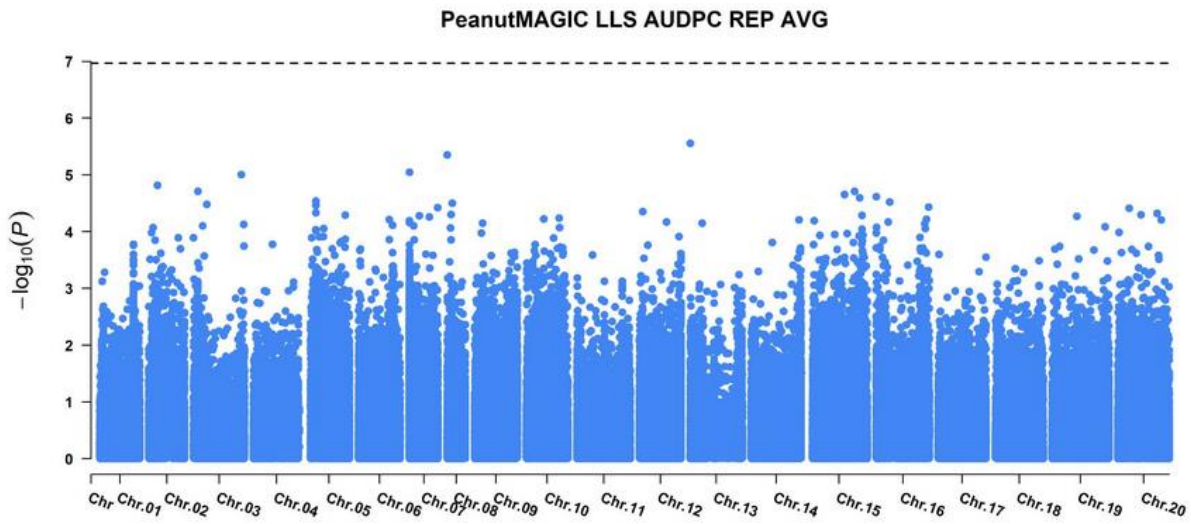


Figure 6.1: Manhattan plot of the average area under the disease progress curve (AUDPC) of late leaf spot (LLS) for all replicates across the three years. No significant associations were found using a Bonferroni corrected *P* value of 0.05.

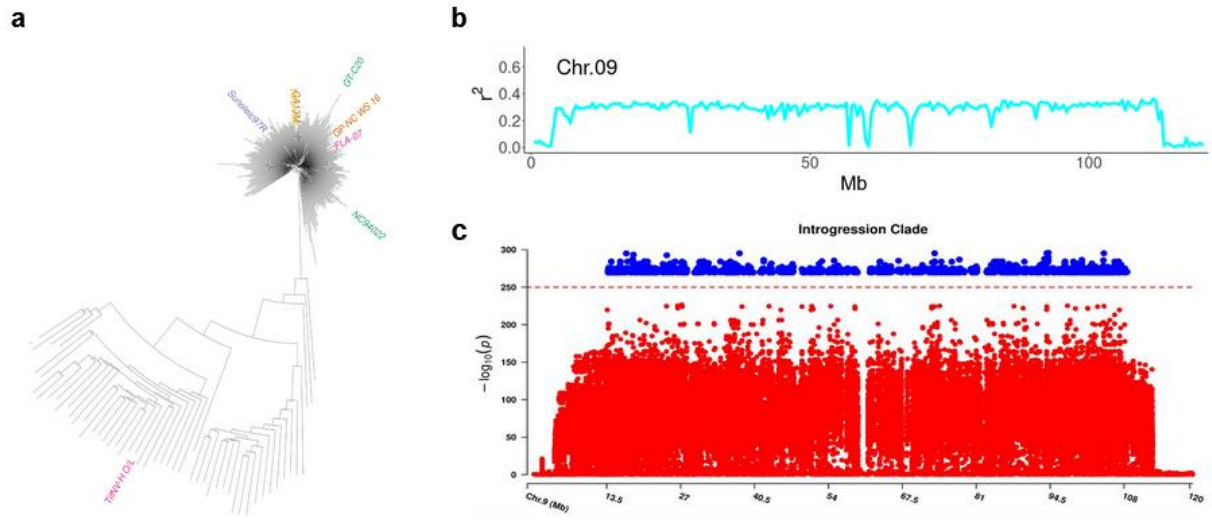


Figure 6.2: Figures from Chapter 2 describing the cause of the population structure of the MAGIC Core using single reference markers. **a**, NJ tree of the MAGIC Core lines and founders using all SNP markers. There were 52 lines and the donor parent TifNV-H O/L as outliers, representing 16.8% of the population. **b**, The r^2 values along Chr. 09, showing higher r^2 values (plateaus) with lower recombination in the regions of *A. cardenasii* introgression, however, there were regions where high recombination rate did occur as denoted by the "valleys" across the r^2 plot. **c**, These locations align with the clades in the phylogenetic tree and the outgroup. The increase in -log (*p* value) (in blue) shows locations the MAGIC Core introgression lines share, where -log (*p* value) was lower indicating the clade with specific introgression recombination points (in red).

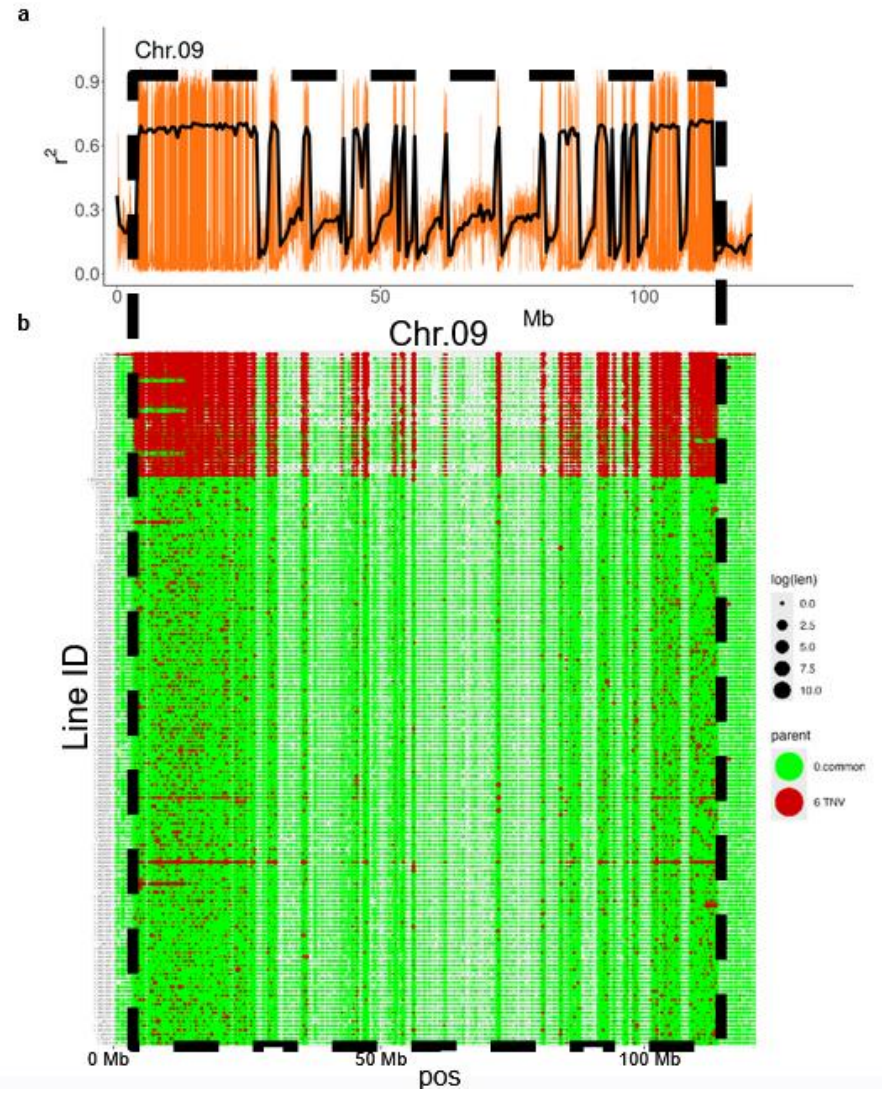


Fig. 6.3: Characteristics of Chr.09 recombination. **a**, r^2 values across Chr.09 calculated from a sliding window of 300 markers depicting select regions with elevated LD. The black line represents the average r^2 value in 0.5 Mb. **b**, PeanutMAGIC Core pangenome marker calls for Chr.09 with founder origin coloration for TNV only. The black dashed box highlights the region of the *A. cardenasii* introgression.

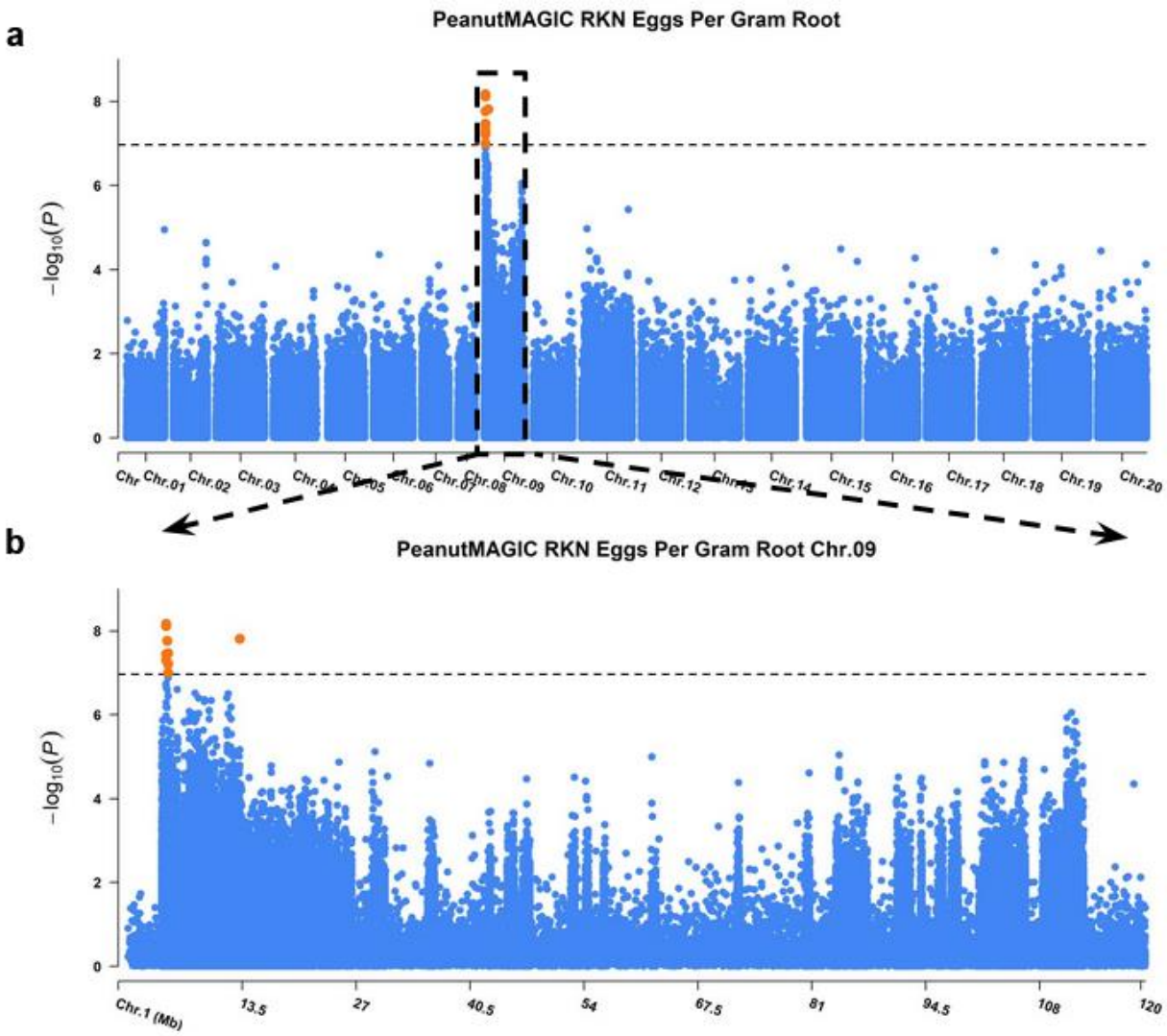


Figure 6.4: **a**, Manhattan plot showing PeanutMAGIC pangenome based markers associated with square root of eggs per gram of root phenotypes. A Bonferroni-corrected *P* value of 0.001 was used as a significant threshold ($P = 2.15 \times 10^{-9}$), represented by a horizontal dashed black line. **b**, Local association plot of Chr.09 showing significant spanning 400 Kb.