



Genome-Wide Association Study of Major Agronomic Traits Related to Domestication in Peanut

Xingguo Zhang, Jianhang Zhang, Xiaoyan He, Yun Wang, Xingli Ma and Dongmei Yin*

College of Agronomy, Henan Agricultural University, Zhengzhou, China

Peanut (*Arachis hypogaea*) consists of two subspecies, *hypogaea* and *fastigiata*, and has been cultivated worldwide for hundreds of years. Here, 158 peanut accessions were selected to dissect the molecular footprint of agronomic traits related to domestication using specific-locus amplified fragment sequencing (SLAF-seq method). Then, a total of 17,338 high-quality single nucleotide polymorphisms (SNPs) in the whole peanut genome were revealed. Eleven agronomic traits in 158 peanut accessions were subsequently analyzed using genome-wide association studies (GWAS). Candidate genes responsible for corresponding traits were then analyzed in genomic regions surrounding the peak SNPs, and 1,429 genes were found within 200 kb windows centered on GWAS-identified peak SNPs related to domestication. Highly differentiated genomic regions were observed between *hypogaea* and *fastigiata* accessions using F_{ST} values and sequence diversity (π) ratios. Among the 1,429 genes, 662 were located on chromosome A3, suggesting the presence of major selective sweeps caused by artificial selection during long domestication. These findings provide a promising insight into the complicated genetic architecture of domestication-related traits in peanut, and reveal whole-genome SNP markers of beneficial candidate genes for marker-assisted selection (MAS) in future breeding programs.

Keywords: peanut, domestication, genome-wide association studies, selective sweeps, single-nucleotide polymorphisms (SNPs)

OPEN ACCESS

Edited by:

Chengdao Li,
Murdoch University, Australia

Reviewed by:

Matthew Nicholas Nelson,
Royal Botanic Gardens, Kew,
United Kingdom
Alice Hayward,
The University of Queensland,
Australia

*Correspondence:

Dongmei Yin
yindm@126.com

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 04 May 2017

Accepted: 04 September 2017

Published: 26 September 2017

Citation:

Zhang X, Zhang J, He X, Wang Y,
Ma X and Yin D (2017) Genome-Wide
Association Study of Major Agronomic
Traits Related to Domestication in
Peanut. *Front. Plant Sci.* 8:1611.
doi: 10.3389/fpls.2017.01611

INTRODUCTION

Peanut, also known as groundnut (*Arachis hypogaea* L.), is one of the most important edible oil crops in the world. Cultivated peanut is an allotetraploid (AABB, $2n = 40$), harboring homoeologous A and B genomes putatively derived from the natural hybridization of two wild diploid species, *A. duranensis* (AA, $2n = 20$) and *A. ipaensis* (BB, $2n = 20$) (Seijo et al., 2007; Moretzsohn et al., 2012). Based on the presence or absence of floral axes on the main stem, cultivated peanut is classified into two subspecies: *hypogaea* and *fastigiata*. Subspecies *hypogaea* is generally described as having a prostrate growth habit with no floral axes on the main stem, while in subspecies *fastigiata* flowers arise on leaf axils on branches as well as the main stem (Krapovickas and Gregory, 1994).

As the only cultivated species of *Arachis*, peanut has been grown for hundreds of years in more than 100 countries worldwide (Huang et al., 2012). The evolution of *Arachis* is therefore closely related to the domestication of cultivated peanut, with a vast number of morphological forms having evolved under cultivation. For example, selection of a more upright growth habit

and shorter branches, which allow easier harvesting and increased seed size, has also resulted in a decrease in resistance to a number of important pathogens (Stalker and Simpson, 1995; Stalker et al., 2013). Domestication-related quantitative trait loci (QTLs) associated with agronomic traits and resistance have already been mapped; however, the utilization of potential alleles has been relatively limited because of the lack of appropriate molecular tools for analysis of these traits in cultivated peanut (Burow et al., 2001; Chu et al., 2011; Ravi et al., 2011; Fonceka et al., 2012; Tseng et al., 2016; Zhou et al., 2016).

With the development of high throughput sequencing technologies, whole-genome sequencing (WGS) has become much more straightforward, allowing analysis of the impact of domestication on genomic variation. Specific-locus amplified fragment sequencing (SLAF-seq) is an efficient method of large-scale single nucleotide polymorphism (SNP) identification and genotyping using high-throughput sequencing, with many advantages such as lower costs and reduced genome complexity (Mamanova et al., 2010; Sun et al., 2013). So far, this new method has been successfully used to address fundamental questions regarding soybean domestication (Han et al., 2016).

Genome-wide association studies (GWAS) have also been used to determine the genetic basis of traits underlying domestication in a wide range of organisms (Lin et al., 2014; Han et al., 2016). However, information on peanut domestication remains limited, largely due to the relatively large size (~2.8 Gb) and complexity of the tetraploid peanut genome (Bertioli et al., 2014). However, in 2014, genomes of the two wild progenitors of cultivated peanut were released by the International Peanut Genome Initiative (IPGI), benefiting studies of agronomic traits related to domestication (Bertioli et al., 2016).

In the present study, high quality SNPs distributed throughout the peanut genome were mined using SLAF-seq of 158 peanut accessions. GWAS was subsequently conducted to identify the genetic architecture of 11 major agronomic traits related to domestication. The results present the first comprehensive view of genome-wide sequence variation in a diverse group of peanut accessions. Moreover, the SNPs and candidate genes related to major agronomic traits will help accelerate peanut breeding programs.

MATERIALS AND METHODS

Plant Materials and Trait Analyses

A total of 158 peanut (*A. hypogaea* L.) accessions were examined in the present study, including 36 *hypogaea* type (group I) with no floral axes on the main stem, and 122 *fastigiata* type (group II) with flowers growing on both branch and the main stem (Table S1). These accessions were elite cultivars collected from different provinces of China, and some of them were greatly produced in special areas.

Seeds from a single plant of each of the 158 accessions were grown in a randomized complete block design with three replications. Four plants from each replicate were then selected to investigate the following 11 agronomic traits: height of the main stem, total number of branches, branch type, leaf color,

pod length, pod width, seed length, seed width, 10-pod weight, 10-seed weight and seed coat color.

SLAF Sample Preparation and Sequencing

Genomic DNA was isolated from fresh leaves of a single plant per accession, and analyzed using SLAF-seq (Sun et al., 2013). To obtain >200,000 SLAF tags per genome, evenly distributed in unique genomic regions, different restriction enzyme combinations were tested using *in silico* digestion prediction. Two restriction enzymes (*RsaI* and *HaeIII*) were selected based on uniqueness and uniformity of simulated fragment alignments to the reference genome sequence of two diploids, *A. ipaensis* (ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000816755.1_Araip1.0, gene model is prefixed by Araip) and *A. duranensis* (ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000817695.1_Aradu1.0, gene model is prefixed by Aradu). Different length fragments of genomic DNA after digestion were then simulated *in silico*. *Oryza sativa indica* (<http://rapdb.dna.affrc.go.jp/>) was selected as the control genome to test the accuracy of the restriction enzyme digestion protocol using SOAP software (Li R. et al., 2009).

A total of 10-ug of genomic DNA from each accession was used for the restriction reaction and subsequent restriction-ligation reactions, including the addition of A to the 3' end and ligation with the Dual-index adapter. PCR was performed with the restriction-ligation samples (diluted) then the PCR products were purified with a Quick Spin column (Qiagen, Hilden, Germany) and electrophoresed on 2% (w/v) agarose gel. Fragments with expected lengths were isolated using a Gel Extraction Kit (Qiagen) and diluted for sequencing. Fragments of 314–344 bp were isolated for use as SLAF tags.

SNP Calling

All reads were processed for quality control and filtered using Seqtk (<https://github.com/lh3/seqtk>). High quality paired-end reads were mapped onto the reference genome (*A. ipaensis* and *A. duranensis*) using the Burrows-Wheeler Aligner (BWA) (Li and Durbin, 2009). Realigner Target Creator and InDel-Realigner in GATK (McKenna et al., 2010) were used to realign InDels, and Unified Genotyper was used to call genotypes across the 158 accessions using the default parameters. Sequencing depths of each sample were calculated using the “Depth of Coverage” module of GATK. Single SNP markers were confirmed using GATK (McKenna et al., 2010) and SAMtools (Li H. et al., 2009). Given the allotetraploid nature of the peanut genome, the genotyping errors caused by partial homologous alignment were resolved by comparing the sequencing depth first, then filtered those SNPs with integrity (genotyped rate) and minor allele frequency (MAF). The exceptionally high homologous regions were not under special analyses, because there were not too much SNPs were found in these regions.

Population Structure Analysis and Phylogenetic Tree Construction

Population structure was calculated using ADMIXTURE software (Alexander et al., 2009). The number of genetic clusters (K) was predefined as 1–10 to explore the population structure

of the tested accessions. This analysis provided maximum likelihood estimates of the proportion of each sample derived from each of the K populations. SNPs were then used to calculate genetic distances among the 158 accessions, and phylogenetic trees were constructed using the neighbor-joining method in MEGA5 (Tamura et al., 2011). Principal component analysis (PCA) was performed using GAPIT software (Lipka et al., 2012).

GWAS of Agronomic Traits

High-integrity SNPs from the tested peanut accessions were used in association analyses using the general linear model (GLM) and compressed mixed linear model (MLM) with TASSEL software (Bradbury et al., 2007). The following formula was used:

$$Y = Xa + Qb + Ku + e$$

where Q is the population structure derived from ADMIXTURE software (Alexander et al., 2009), K is the relationship between samples obtained from SPAGeDi (Hardy and Vekemans, 2002), using Q in GLM and Q + K in MLM. X represents the genotype and Y the phenotype, allowing associated values of each SNP to be calculated. A value of <math> < 0.01 </math> was used as the threshold to determine the existence of a significant association. Gene predictions were annotated according to the method used in Zhang et al. (2015). Candidate genes associated with each trait located within a 100-kb region upstream or downstream of peak SNPs, because the size of the larger linkage disequilibrium (LD) Block is mostly distributed around 200 kb when the whole genome was analyzed for LD Block. The r^2 value (Marker_Rsq, is the marginal R-squared for the marker) was used to explain the phenotypic variation of each marker. It was calculated as SS Marker (after fitting all other model terms) / SS Total, where SS stands for sum of squares (<https://bitbucket.org/tasseladmin/tassel-5-source/wiki/UserManual/GLM/GLM>).

Population Differentiation (F_{ST}) and Putative Selective Sweeps

The divergence index, F -statistics (F_{ST}), is a measure of population differentiation or genetic distance based on genetic polymorphism data (Hudson et al., 1992). To determine potentially differentiated regions, F_{ST} estimations and sequence diversity (π) ratios were evaluated using a 100-kb sliding window with 10-kb steps (Lam et al., 2010). Highly differentiated genomic regions with a significant F_{ST} value ($p = 5\%$) and the top 5% of π ratios were defined as potential selective sweeps (Li et al., 2013).

RESULTS

Phenotypic Variation among Peanut Accessions

Phenotypic evaluation revealed a broad range of variation among the 158 peanut accessions (Figure 1, Table S1). The descriptive statistics of phenotypic variation of eight traits were listed in Table S2. Height of main stem ranged from 12 to 87.33 cm, with an average of 38.35 cm. Total number of branches also varied with an average of 12.8. All accessions showed continuous distribution of the three pod-related and three seed-related traits, with a

coefficient of variation (CV) of 18.63% for seed length and 66.98% for the total number of branches, suggesting a quantitative inheritance pattern. The remaining three traits, branch type, leaf color and seed coat color, were evaluated as quality variation.

SLAF-Tags and SNP Data

A total of 369,725 high quality SLAFs evenly distributed on 10 A and 10 B chromosomes were obtained from 397.19 M paired-end reads after sequence alignment with the reference genome (*A. duranensis* and *A. ipaensis*). Sequencing depths ranged from 32.42 to 4.94 X, with an average of 8.06 X. Polymorphic SLAFs defined by both GATK and SAMtools were recorded as reliable SNPs, resulting in a total of 268,889 SNP markers among the 158 accessions. After filtering SNPs located on scaffolds, 17,338 high quality SNPs with an MAF > 0.05 and integrity > 0.5 were selected for further analyses (Table 1, Figure S1).

The selected SNP markers were not evenly distributed across the whole genome, with 7,538 and 9,800 markers on the A and B subgenomes, respectively. Chromosome B03 harbored the highest proportion of SNPs (8.23%; 1,427 of 17,338), while chromosome A08, the shortest chromosome at 48.94 Mb, contained the least (1.81%; 314 of 17,338). The average number of SNPs/Mb was seven on both the A and B subgenomes, while average genes/Mb were 35 and 31, respectively. Chromosome B03 had the highest number of genes (5,188 of 77,617), while its counterpart, chromosome A03, contained the second highest (4,929 of 77,617).

LD was estimated as the r^2 value, revealing uneven distribution of SNPs on each chromosome. r^2 values of the A subgenome ranged from 0.071 on chromosome A03 to 0.356 on chromosome A02, while those of the B subgenome ranged from 0.029 on chromosome B03 to 0.251 on chromosome B05. Average r^2 values of the A and B subgenomes were 0.177 and 0.137, respectively, revealing differences in the level of LD between different chromosomes and subgenomes.

Population Structure and Genome-Wide Divergence in Peanut

To examine divergence of the 158 accessions during evolution, analysis of population structure, phylogenetic relationships and PCA were carried out using the 17,338 selected SNPs. According to the K genetic clusters, the most likely number of inferred members was 2 with $\Delta K = 0.47$ (Figure 2C). Nevertheless, accessions were classified into three major clusters in the NJ phylogenetic tree, with *hypogaea* accessions (group I) forming a separate cluster, and *fastigiata* accessions (group II) forming another large cluster (Figure 2B). PCA was conducted using the first two principal components (Figure 2A), PC1 (with variance explain 8.16%) and PC2 (with variance explain 5.23%), revealing that the accessions probably divided into two groups, with different degrees of introgression between the two subspecies during cultivation.

GWAS of Loci Underlying Domestication Traits in Peanut

Eleven traits were selected for identification of underlying genetic loci and regions. In total, 51 SNP peaks associated with six

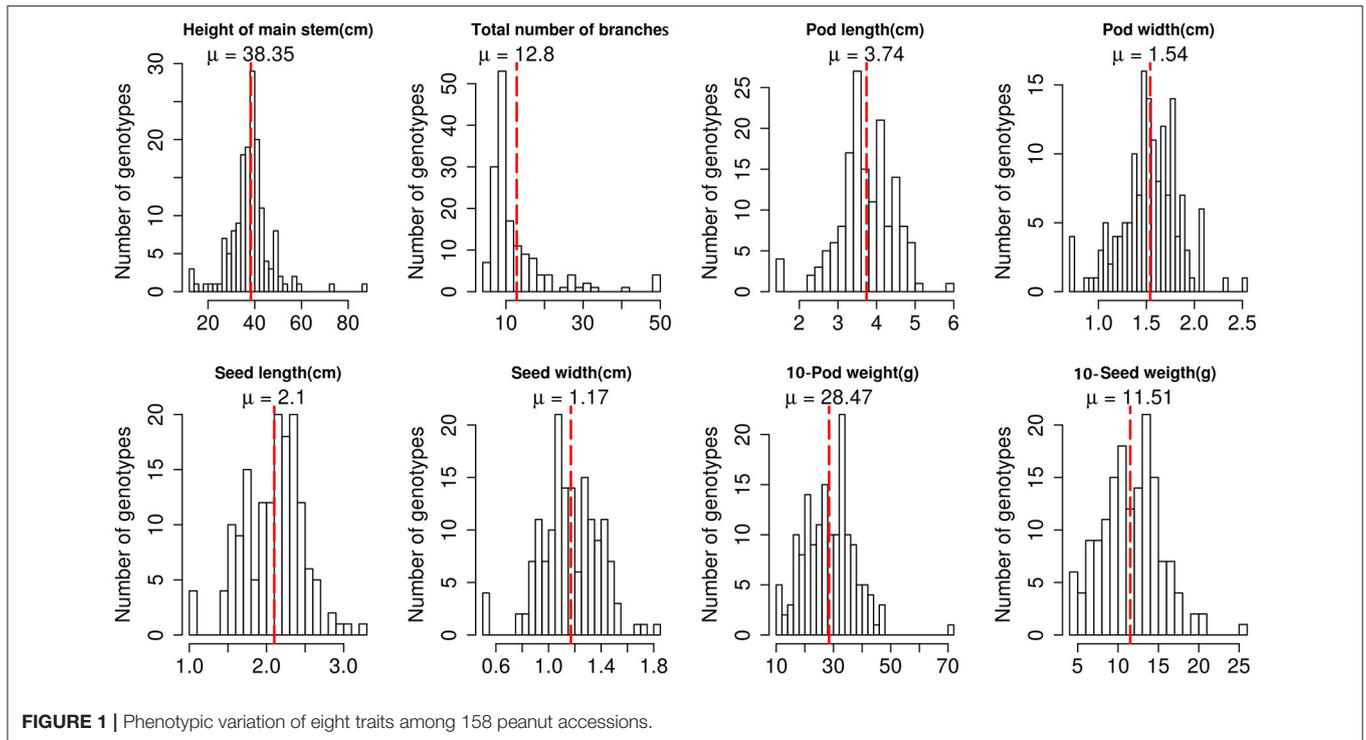
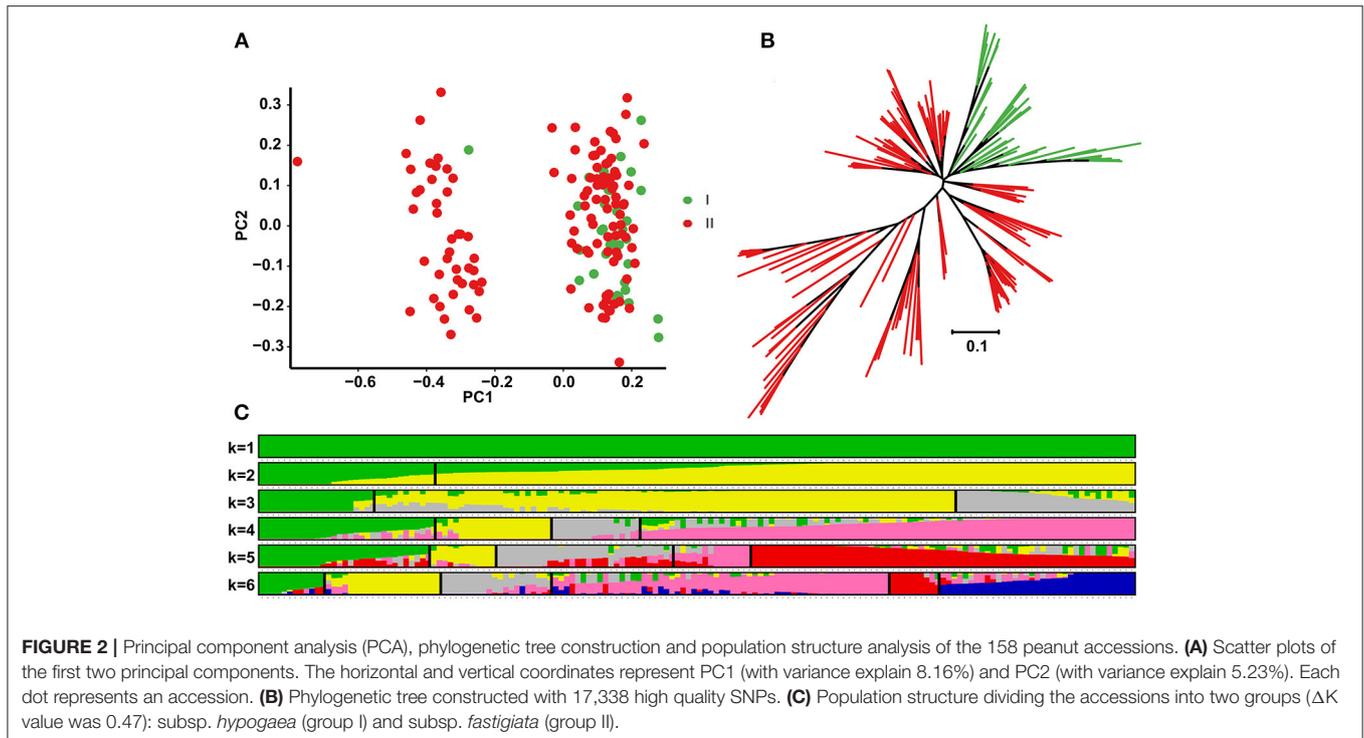


TABLE 1 | Distribution of 17,338 SNPs in 20 chromosomes identified in 158 peanut accessions.

Chr	Chr length(Mb)	Number of SNPs	Number of genes	Average number of SNPs per Mb	Average number of genes per Mb	r^2 of chr LD
A01	106.02	725	3493	7	33	0.177
A02	92.63	1059	3129	11	34	0.356
A03	133.14	970	4929	7	37	0.071
A04	121.18	749	3833	6	32	0.185
A05	108.28	793	3868	7	36	0.166
A06	110.73	862	3712	8	34	0.149
A07	77.95	482	2850	6	37	0.161
A08	48.94	314	2940	6	60	0.151
A09	119.00	887	3694	7	31	0.173
A10	107.26	697	3488	6	33	0.186
B01	136.91	1073	3791	8	28	0.084
B02	108.64	710	3580	7	33	0.217
B03	135.60	1427	5133	11	38	0.029
B04	133.21	1047	4276	8	32	0.159
B05	149.44	1011	4433	7	30	0.251
B06	136.72	876	4372	6	32	0.094
B07	125.99	1093	3813	9	30	0.032
B08	129.15	962	3710	7	29	0.074
B09	146.50	654	4409	4	30	0.196
B10	135.81	947	4164	7	31	0.238
A	1025.12	7538	35936	7	35	0.177
B	1337.96	9800	41681	7	31	0.137
AB	2363.08	17338	77617	7	33	0.157

Chr, Chromosome; r^2 of chr LD, average LD between all SNPs on a chromosome.



traits reached the corrected P value according to the Bonferroni method ($P < 5.76e^{-07}$ at $\alpha = 0.01$ or $-\log_{10}(P) = 6.238$) (Figure 3, Table S3). A 100-kb genomic region on each side of the peak SNP associated with the corresponding traits was subsequently analyzed for identification of candidate genes (Table 2).

As a result, a total of 13 significant SNPs were associated with height of the main stem, with the peak SNP A03-26481539 explaining 27.55% of the phenotypic variation. The Aradu 52T5J gene was located ~ 32 kb from this SNP, and is known to encode a malate dehydrogenase, which is thought to be related to biomass and plant height in maize (Carrari et al., 2005). Peak SNP A10-90376017 for the total number of branches explained 26.64% of the phenotypic variation, and was located only ~ 8 kb from the nearest gene, Aradu J85DC, which encodes a cytochrome P450 superfamily protein reportedly involved in the strigolactone synthetic pathway in rice (Gomez-Roldan et al., 2008) and soybean. The peak SNP A03-6992035 explained 15.52 and 19.99% of the phenotypic variation of seed length and 10-seed weight, respectively. A candidate gene encoding a bHLH transcription factor was located ~ 57 kb from this peak, and is thought to be a pleiotropic gene involved in seed development (Kondou et al., 2008). Aradu PZ2UH was located ~ 42 kb from A03-119879303, another peak SNP related to 10-seed weight (explaining 18.69% phenotypic variation) on chromosome A03, and encodes an auxin response factor (ARF) involved in plant growth and seed development (Okushima et al., 2005; Attia et al., 2009). A peak SNP explaining 26.32% of the phenotypic variation in pod width was located on chromosome A05. One candidate gene, Aradu CVC5Q, was located ~ 32 kb away, and encodes a microtubule-associated protein (MAP)

that reportedly influences seed shape by regulating microtubule growth (Deng et al., 2012). The candidate gene for seed coat color, located ~ 12 kb from the peak SNP B03-22076736, which explained 21.94% of the phenotypic variation), also represented a bHLH transcription factor previously found to influence seed coat color in rice (Sweeney et al., 2006). These results suggest that GWAS was effective in clarifying candidate genes related to major domestication-related traits in peanut.

Genomic Changes and Target Regions Associated with Selection

Genomic changes related to selective processes during domestication can be determined using genotypic data. In this study, a total of 1,429 genes were defined in 335 highly differentiated genomic regions in the two groups using an F_{ST} threshold of 0.261 (determined by the 5% right tails of the F_{ST} distribution) and a $\pi I / \pi II$ ratio threshold of 2.03 (Figures 4B,C). The gene distribution of the resulting sweeps was subsequently determined (Table S4). A03 contained the most number of genes (662 genes, 45.94%), and presented stronger selective sweep signals in group I than group II. A total of 186 and 158 genes were found on chromosomes B03 and B08, respectively, corresponding to 12.91 and 11.06%, respectively, and ranking them second and third during peanut artificial selection. No selective sweep signals were detected on chromosomes B01 or B10.

A total of 15 SNPs were found in the 200-kb selective sweep regions of major peak SNP A03-6992035 (Figure 4A), which is related to seed length and seed weight. Three SNPs were also found in the two gene models, Aradu D69CU and Aradu T1PSR,

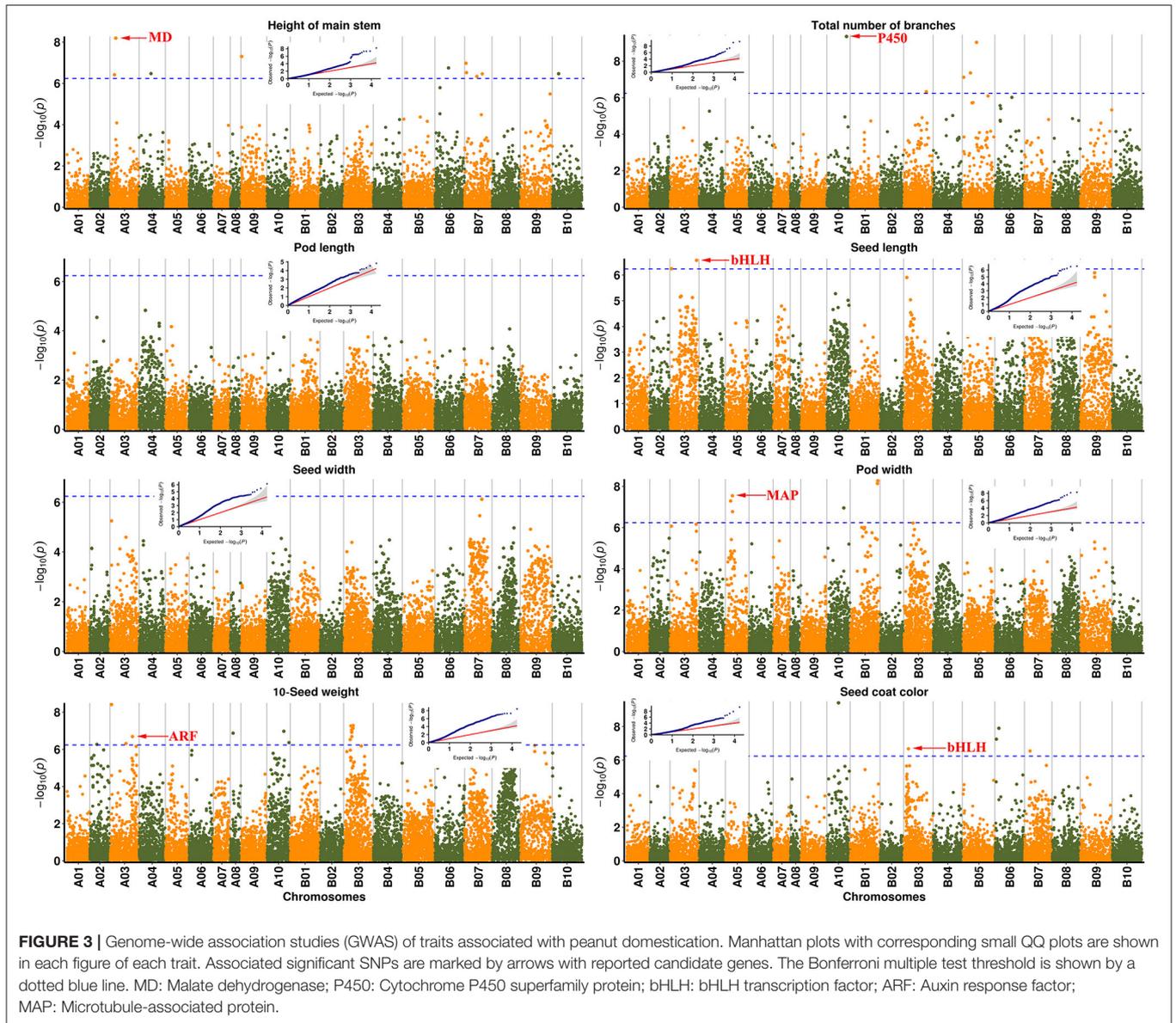


FIGURE 3 | Genome-wide association studies (GWAS) of traits associated with peanut domestication. Manhattan plots with corresponding small QQ plots are shown in each figure of each trait. Associated significant SNPs are marked by arrows with reported candidate genes. The Bonferroni multiple test threshold is shown by a dotted blue line. MD: Malate dehydrogenase; P450: Cytochrome P450 superfamily protein; bHLH: bHLH transcription factor; ARF: Auxin response factor; MAP: Microtubule-associated protein.

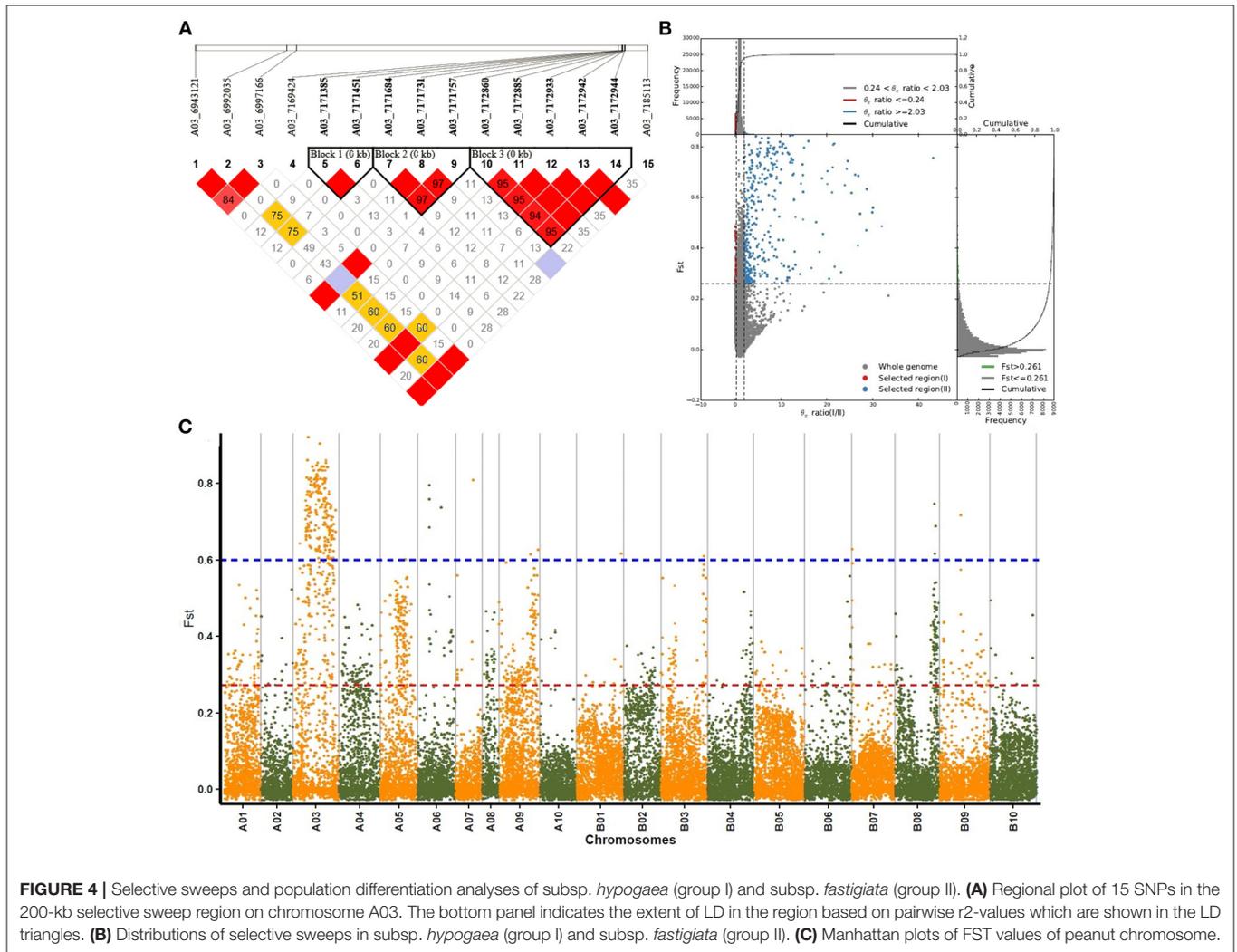
TABLE 2 | Six significant SNPs and predicted candidate genes associated with major agronomic traits in 158 peanut accessions.

Trait	SNP location (bp)	Gene model	Distance to SNP (kb)	Functional annotation
HMS	A03-26481539	Aradu 52T5J	32 (26,514,081–26,516,597)	Malate dehydrogenase
TNB	A10-90376017	Aradu J85DC	8 (90,384,547–90,391,660)	Cytochrome P450 superfamily protein
Seed length	A03-6992035	Aradu T1PSR	57 (7,049,674–7,051,562)	bHLH transcription factor
Seed weight	A03-119879303	Aradu PZ2UH	42 (1,19,921,885–1,19,926,789)	Auxin response factor
Pod weight	A05-32373760	Aradu CVC5Q	32 (32,341,000–32,343,951)	Microtubule-associated protein
Seed coat color	B03-22076736	Araip P4GTD	12 (22,064,568–22,066,022)	bHLH transcription factor

HMS, Height of main stem; TNB, Total number of branches.

respectively, both of which encode a bHLH transcription factor. In total, 21 genes were found in this region, and only nine showed annotation. In addition to the two bHLH genes mentioned

above, three major intrinsic protein (MIP) genes involved in carbohydrate transport and metabolism, and one F-box gene and one proline-rich protein (PRP) gene both involved in



plant growth and stress response, were also identified in this region.

Nucleotide-binding leucine-rich repeat (NB-LRR)-encoding genes are of particular interest because they confer resistance against pests and disease. Possible resistant genes in the highly differentiated genomic regions were therefore analyzed. Ten and 25 genes containing the NB-LRR domains were identified on the A and B subgenomes, respectively. Of the 10 NB-LRR-encoding genes on the A subgenome, eight were located on chromosome A03, and of the 25 on the B subgenome, 18 were located in two genomic regions of chromosome B07. One region was located within 282 kb of chromosome B07 (between 1714315 and 1996823) and harbored 15 NB-LRR-encoding genes (Table S5), suggested that this area contains a resistance gene family. The second region was located between 4775007 and 4820955 on chromosome B07, and included three NB-LRR-encoding genes. Interestingly, five NB-LRR-encoding genes were found in a major selective sweep on chromosome B03. This region covered 6.89 Mb and contained 107 selective genes, including one gene encoding a

Gibberellin-related protein and three genes related to flavonoid biogenesis and regulation. These findings suggest that this region plays important roles in both resistance and plant-type-related traits.

DISCUSSION

Species in the genus *Arachis* are widely distributed across tropical, subtropical and warm temperate zones, but only the cultivated peanut (*A. hypogaea*) is an important food crop. Peanut evolved morphologically during domestication, allowing it to adapt to various agroecological environments (Stalker and Simpson, 1995). Subsp. *fastigiata* has more advanced traits than subsp. *hypogaea* in terms of plant habit and pod morphology, and in this study, represented a larger portion (77.2%) of the selected accessions (Figure 1 and Table S1). In line with this, Krapovickas (1969) postulated that *hypogaea* (subsp. *hypogaea*) represents the most ancient variety due to its runner habit, lack of floral spikes and branching patterns, which are similar to the characteristics of wild *Arachis* species. In this study,

analysis of phylogenetic relationships, population structure and PCA among 158 peanut accessions revealed that accessions in subsp. *fastigiata* contain unique genomic regions that differ from those in *hypogaea* (Figure 2). Four wild diploid accessions were previously sequenced and removed due to ploidy difference (data not shown). Interestingly, some accessions were not clearly distinguishable, possibly because they underwent differing degrees of genetic introgression during manual selection.

Cultivated peanut has a narrow genetic base, possibly resulting from a single polyploidization event (Kochert et al., 1996), and can therefore be improved using introgression genes for disease resistance and other important agronomic traits from wild species (Dwivedi et al., 2007; Holbrook et al., 2008; Isleib et al., 2011). Molecular markers for economically significant traits have been widely used to improve the speed and efficiency of MAS breeding in peanut (Selvaraj et al., 2009; Chu et al., 2011). SNP markers, the most abundant molecular marker, are a cost-effective method of high-throughput molecular genotyping, but have limited use in peanut because of the homeologous A and B subgenomes. In this study, the two diploid genome (*A. ipaensis* and *A. duranensis*) were used as the reference genomes. In order to accurately allocate marker tags to the correct genome, sequencing depth was combined with integrity and MAF to filter SNPs. There are not too much SNPs found in exceptionally high homologous regions, which would not influence the following GWAS results. Finally, a total of 17,338 high quality SNPs were identified across the whole genome by reduced representation sequencing technology (the SLAF-seq method) (Table 1 and Figure S1). The average sequencing depth was 8.06-fold, and the average number of SNPs / Mb was seven in both the A and B subgenomes, suggesting a reasonable density of markers given the relatively low cost of the genotyping method (He et al., 2011; Li et al., 2013; Morris et al., 2013).

GWAS is considered an efficient method for genetic analysis of complex trait variation (Han et al., 2016; Fahrenkrog et al., 2017). Based on the markers developed in this study, 51 association SNP peaks were identified in this study, along with candidate genes or loci corresponding to domestication-related traits (Figure 3 and Table S3). A total of 13 association peaks for height of the main stem were found on eight chromosomes, while 18 peaks representing seed weight were found on six chromosomes, suggesting that there is a group of QTLs responsible for these traits. The basic transcription factor, bHLH family, is involved in plant growth and developmental processes as well as stress responses and secondary metabolism (Heim et al., 2003), and in this study, was linked with seed shape and seed coat color (Table 2). The Aradu PBR53 gene located ~37 kb from SNP A05-32373760 for pod weight, which was predicted to encode formin protein, a newly revealed regulatory factor of cell skeleton assembly (Guo and Ren, 2006). In *Arabidopsis thaliana*, it was found to play a role in cell division and cell polarity (Zhang et al., 2016), but studies in peanut are limited, making it a good focal point for future research. Using the SNPs identified in this study, further analysis of agronomic traits will be possible, allowing rapid identification of candidate genes for future peanut breeding programs.

Similar to other important crop plants, peanut has undergone continuous selection through domestication and intensive breeding events. Domestication from a wild to a cultivated species is largely associated with genome-wide duplications, mutations, selection and genetic drift (Kim et al., 2010; Li et al., 2013). Many traits thought to be involved in peanut domestication were previously clustered in three genomic regions on chromosomes A07, B02, and B05 (Fonckea et al., 2012), suggesting that several linked genes are responsible for the phenotypic variation in 158 peanut accessions collected from all growing regions of China. In this study, selective sweeps in the two peanut groups were measured using F_{ST} values and sequence diversity (π) ratios. The existence of major selective sweeps on chromosome A03 indicated that this chromosome was subjected to primary selection pressure (Figure 4). This finding was similar to the results of GWAS, whereby several important genes related to domestication traits were located on chromosome A03. SNPs in these regions are therefore likely to be valuable for MAS breeding.

There are several major diseases, like early leaf spot (caused by *Cercospora arachidicola*), late leaf spot (caused by *Cercosporidium personatum*), and Tomato spotted wilt virus (TSWV, spread by thrips), may cause significant yield loss in peanut production (Nigam et al., 2012). Breeding of resistance cultivars is the most cost-effective method of reducing disease damage in peanuts. The role of NB-LRR proteins in plant defense against pathogens has been extensively studied (DeYoung and Innes, 2006; Nagy and Bennetzen, 2008). In this study, genes encoding NB-LRR protein were mainly distributed on chromosomes A03 and B07, especially in a 280 kb region on B07, which contained 15 promising candidate genes (Table S5). Wang et al. (2013) constructed two genetic maps to identify QTLs for thrips, TSWV and leaf spot (LS) in peanut. One QTL for TSWV, qF₂TSWV3, was identified in the same marker interval (seq5D5-GM2744) on linkage group AhII with one QTL for LS, qF₂LS1. Another QTL, qF₅LS10 for LS, was identified between GM1254 and seq15C10 on linkage LGT17. The two linkage groups, AhII and LGT17, were colinearized with B07 of reference consensus genetic map (the three QTLs mentioned above covered 24 cM), suggested that there is a cluster of resistance QTLs on B07. Pandey et al. (2017) found that of the total 42 QTLs linked to diseases resistance in peanut, 34 were mapped on the A sub-genome and eight mapped on the B sub-genome, suggesting that the A sub-genome harbors more resistance genes than the B sub-genome, which is in agreement with Bertoli et al. (2016), who reported that there are more NB-LRR-encoding disease resistance-like genes in the “A” genome (397 genes in *A. duranensis*) than in the “B” genome (345 genes in *A. ipaensis*). Though function of these genes needs to be further validated, these findings suggest that SNPs located in these major selective sweeps will facilitate future breeding of resistant cultivars in peanut. The major selective sweep on chromosome B03 harbored a number of genes related to both resistance and plant-type-related traits, similar to a recent study (Zhou et al., 2016), suggesting that this genetic region is worthy of further investigation.

CONCLUSION

In summary, this study provides the first insight into the complex genetic relationship between agronomic traits and domestication processes in peanut. Chromosomes A03 and B03 harbored major genes related to peanut domestication, while chromosome B07 contained a cluster of NB-LRR-encoding genes. Further studies are now needed to understand the genetic mechanisms underlying yield- and seed-related traits and identify potential resistant genes for future peanut breeding programs.

AUTHOR CONTRIBUTIONS

XZ, JZ, XH, DY carried out phenotyping and genotyping. XZ, YW, DY managed the project. XZ, XM, DY analyzed the data.

REFERENCES

- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109
- Attia, K. A., Abdelkhalik, A. F., Ammar, M. H., Wei, C., Yang, J. S., Lightfoot, D. A., et al. (2009). Antisense phenotypes reveal a functional expression of *OsARF1*, an auxin response factor, in transgenic rice. *Curr. Issues Mol. Biol.* 11(Suppl. 1), 29–34. doi: 10.21775/cimb.011.i29
- Bertioli, D. J., Cannon, S. B., Froenicke, L., Huang, G., Farmer, A. D., Cannon, E. K. S., et al. (2016). The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat. Genet.* 48, 438–446. doi: 10.1038/ng.3517
- Bertioli, D. J., Ozias-Akins, P., Chu, Y., Dantas, K. M., Santos, S. P., Gouvea, E., et al. (2014). The use of SNP markers for linkage mapping in diploid and tetraploid peanuts. *Genes Genome Genet.* 4, 89–96. doi: 10.1534/g3.113.007617
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308
- Burrow, M. D., Simpson, C. E., Starr, J. L., and 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, 823–837.
- Carrari, F., Loureiro, M. E., Ratcliffe, R. G., Sweetlove, L. J., and Fernie, A. R. (2005). Enhanced photosynthetic performance and growth as a consequence of decreasing mitochondrial malate dehydrogenase activity in transgenic tomato plants. *Plant Physiol.* 137, 611–622. doi: 10.1104/pp.104.055566
- Chu, Y., Wu, C. L., Holbrook, C. C., Tillman, B. L., Person, G., and Ozias-Akins, P. (2011). Marker-assisted selection to pyramid nematode resistance and the high oleic trait in peanut. *Plant Genome* 4, 110–117. doi: 10.3835/plantgenome2011.01.0001
- Deng, Z. Y., Liu, L. T., Li, T., Yan, C. J., and Wang, T. (2012). “SAR1 protein affect grain size and shape by regulating cell microtubule depolymerization in rice,” in *From the Plant Science to the Agricultural Development: Symposium of National Congress of Plant Biology, NCPB-2012* (Shanxi).
- DeYoung, B. J., and Innes, R. W. (2006). Plant NBS-LRR proteins in pathogen sensing and host defense. *Nat. Immunol.* 7, 1243–1249. doi: 10.1038/ni1410
- Dwivedi, S. L., Bertioli, D. J., Crouch, J. H., Valls, J. F., Upadhyaya, H. D., Fávero, A., et al. (2007). “Peanut” in *Genome Mapping and Molecular Breeding in Plants. Oilseeds, Vol. 2*, ed C. Kole (Berlin; Heidelberg: Springer-Verlag), 115–151.
- Fahrenkrog, A. M., Neves, L. G., Resende, M. F. R. Jr., Vazquez, A. I., Campos, G., Dervinis, C., et al. (2017). Genome-wide association study reveals putative regulators of bioenergy traits in *Populus deltoides*. *N. Phytol.* 213, 799–811. doi: 10.1111/nph.14154
- Fonceca, D., Tossim, H. A., Rivallan, R., Vignes, H., Faye, I., Ndoye, O., et al. (2012). Fostered and left behind alleles in peanut: interspecific QTL mapping

XZ wrote the paper. All of the authors read and approved the manuscript.

ACKNOWLEDGMENTS

This work was financially supported by grants from the National Natural Science Foundation of China (No. 31471525) and key scientific and technological project in Henan Province (No. 161100111000).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.01611/full#supplementary-material>

- reveals footprints of domestication and useful natural variation for breeding. *BMC Plant Biol.* 12:26. doi: 10.1186/1471-2229-12-26
- Gomez-Roldan, V., Femas, S., Brewer, P. B., Puech-Pages, V., Dun, E. A., Pillot, J. P., et al. (2008). Strigolactone inhibition of shoot branching. *Nature* 455, 189–194. doi: 10.1038/nature07271
- Guo, Q. C., and Ren, H. Y. (2006). Formins: bringing new insights to the organization of actin cytoskeleton. *Chinese Sci. Bull.* 51, 2937–2943. doi: 10.1007/s11434-006-2214-x
- Han, Y. P., Zhao, X., Liu, D. Y., Li, Y. H., Lightfoot, D. A., Yang, Z. J., et al. (2016). Domestication footprints anchor genomic regions of agronomic importance in soybeans. *N. Phytol.* 209, 871–884. doi: 10.1111/nph.13626
- Hardy, O. J., and Vekemans, X. (2002). SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Resour.* 2, 618–620. doi: 10.1046/j.1471-8286.2002.00305.x
- He, Z., Zhai, W., Wen, H., Tang, T., Wang, Y., Lu, X., et al. (2011). Two evolutionary histories in the genome of rice: the roles of domestication genes. *PLoS Genet.* 7:e1002100. doi: 10.1371/journal.pgen.1002100
- Heim, M. A., Jakoby, M., Werber, M., Martin, C., Weisshaar, B., and Bailey, P. (2003). The basic helix-loop-helix transcription factor family in plants: A genome-wide study of protein structure and functional diversity. *Mol. Biol. Evol.* 20, 735–747. doi: 10.1093/molbev/msg088
- Holbrook, C. C., Timper, P., Culbreath, A. K., and Kvien, C. K. (2008). Registration of ‘Tifguard’ peanut. *J. Plant Reg.* 2, 391–410. doi: 10.3198/jpr2007.12.0662crc
- Huang, L., Jiang, H. F., Ren, X. P., Chen, Y. N., Xiao, Y. J., Zhao, X. Y., et al. (2012). Abundant microsatellite diversity and oil content in wild *Arachis* species. *PLoS ONE* 7:e50002. doi: 10.1371/journal.pone.0050002
- Hudson, R. R., Slatkin, M., and Maddison, W. P. (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics* 132, 583–589.
- Isleib, T. G., Milla-Lewis, S. R., Pattee, H. E., Copeland, S. C., Zuleta, M. C., Shew, B. B., et al. (2011). Registration of ‘Bailey’ Peanut. *J. Plant Reg.* 5, 27–39. doi: 10.3198/jpr2009.12.0742crc
- Kim, M. Y., Lee, S., Van, K., Kim, T. H., Jeong, S. C., Choi, I. Y., et al. (2010). Whole-genome sequencing and intensive analysis of the undomesticated soybean (*Glycine soja* Sieb. and Zucc.) genome. *Proc. Natl. Acad. Sci. U.S.A.* 107, 22032–22037. doi: 10.1073/pnas.1009526107
- Kochert, G., Stalker, H. T., Gimenes, M., Galgano, L., Romero, L. C., and Moore, K. (1996). RFLP and cytogenetic evidence on the origin and evolution of the allotetraploid domesticated peanut, *Arachis hypogaea* (Leguminosae). *Am. J. Bot.* 83, 1282–1291. doi: 10.2307/2446112
- Kondou, Y., Nakazawa, M., Kawashima, M., Ichikawa, T., Yoshizumi, T., Suzuki, K., et al. (2008). Retarded growth of embryo 1, a new basic helix-loop-helix protein, expresses in endosperm to control embryo growth. *Plant Physiol.* 147, 1924–1935. doi: 10.1104/pp.108.118364
- Krapovickas, A. (1969). “The origin, variability and spread of the groundnut *Arachis hypogaea*; English translation,” in *The Domestication and Exploitation*

- of *Plants and Animals*, eds P. J. Ucko and I. S. Falk (London: Gerald Duckworth Co Ltd.), 424–441.
- Krapovickas, A., and Gregory, W. C. (1994). Taxonomy of the genus *Arachis* (*Leguminosae*). *Bonplandia* 8, 1–186.
- Lam, H. M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F. L., et al. (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* 42, 1053–1059. doi: 10.1038/ng.715
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, M., Tian, S., Jin, L., Zhou, G., Li, Y., Zhang, Y., et al. (2013). Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat. Genet.* 45, 1431–1438. doi: 10.1038/ng.2811
- Li, R., Yu, C., Li, Y., Lam, T. W., Yiu, S. M., and Kristiansen, K. (2009). SOAP 2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25, 1966–1967. doi: 10.1093/bioinformatics/btp336
- Lin, T., Zhu, G. T., Zhang, J. H., Xu, X. Y., Du, Y. C., and Huang, S. W. (2014). Genomic analyses provide insights into the history of tomato breeding. *Nat. Genet.* 46, 1220–1226. doi: 10.1038/ng.3117
- Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPI: genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. doi: 10.1093/bioinformatics/bts444
- Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., et al. (2010). Target-enrichment strategies for next generation sequencing. *Nat. Methods* 7, 111–118. doi: 10.1038/nmeth.1419
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a map reduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Moretzsohn, M. C., Gouvea, E. G., Inglis, P. W., Leal-Bertioli, S. C. M., Valls, J. F. M., and Bertioli, D. J. (2012). A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. *Ann. Bot.* 111, 113–126. doi: 10.1093/aob/mcs237
- Morris, G. P., Ramu, P., Deshpande, S. P., Hash, C. T., Shah, T., Upadhyaya, H. D., et al. (2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl. Acad. Sci. U.S.A.* 110, 453–458. doi: 10.1073/pnas.1215985110
- Nagy, E. D., and Bennetzen, J. L. (2008). Pathogen corruption and site-directed recombination at a plant disease resistance gene cluster. *Genome Res.* 18, 1918–1923. doi: 10.1101/gr.078766.108
- Nigam, S. N., PrasadaRao, R. D. V. J., Bhatnagar-Mathur, P., and Sharma, K. K. (2012). Genetic management of virus diseases in peanut. *Plant Breed. Rev.* 36, 293–356. doi: 10.1002/9781118358566.ch4
- Okushima, Y., Overvoorde, P. J., Arima, K., Alonso, J. M., Chan, A., Chang, C., et al. (2005). Functional genomic analysis of the AUXIN RESPONSE FACTOR gene family members in *Arabidopsis thaliana*: unique and overlapping functions of ARF7 and ARF19. *Plant Cell* 17, 444–463. doi:10.1105/tpc.104.028316
- Pandey, M. K., Wang, H., Khara, P., Vishwakarma, M. K., Kale, S. M., Culbreath, A. K., et al. (2017). Genetic dissection of novel QTLs for resistance to leaf spots and tomato spotted wilt virus in peanut (*Arachis hypogaea* L.). *Front. Plant Sci.* 8:25. doi: 10.3389/fpls.2017.00025
- Ravi, K., Vadez, V., Isobe, S., Mir, R. R., Guo, Y., Nigam, S. N., et al. (2011). Identification of several small main-effect QTLs and a large number of epistatic QTLs for drought tolerance related traits in groundnut (*Arachis hypogaea* L.). *Theor. Appl. Genet.* 122, 1432–2242. doi: 10.1007/s00122-010-1517-0
- Seijo, G., Lavia, G. I., Fernández, A., Krapovickas, A., Ducasse, D. A., Bertioli, D. J., et al. (2007). Genomic relationships between the cultivated peanut (*Arachis hypogaea*, Leguminosae) and its close relatives revealed by double GISH. *Am. J. Bot.* 94, 1963–1971. doi: 10.3732/ajb.94.12.1963
- Selvaraj, M. G., Narayana, M., Schubert, A. M., Ayers, J. L., Baring, M. R., and Burow, M. D. (2009). Identification of QTLs for pod and kernel traits in cultivated peanut by bulked segregant analysis. *Electron J. Biotech.* 12, 3–4. doi: 10.2225/vol12-issue2-fulltext-13
- Stalker, H. T., and Simpson, C. E. (1995). “Germplasm resources in *Arachis*,” in *Advances in Peanut Science*, eds H. E. Pattee and H. Thomas Stalker (Stillwater, OK: American Peanut Research and Education Society), 14–53.
- Stalker, H. T., Tallury, S. P., Ozias-Akins, P., Bertioli, D., and Bertioli, S. C. L. (2013). The value of diploid peanut relatives for breeding and genomics. *Peanut Sci.* 40, 70–88. doi: 10.3146/PS13-6.1
- Sun, X., Liu, D., Zhang, X., Li, W., Liu, H., Hong, W., et al. (2013). SLAF-seq: an efficient method of large-scale de novo SNP discovery and genotyping using high-throughput sequencing. *PLoS ONE* 8:e58700. doi:10.1371/journal.pone.0058700
- Sweeney, M. T., Thomson, M. J., Pfeil, B. E., and McCouch, S. (2006). Caught red-handed: Rc encode a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell* 18, 283–294. doi: 10.1105/tpc.105.038430
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121
- Tseng, Y. C., Tillman, B. L., Peng, Z., and Wang, J. P. (2016). Identification of major QTLs underlying tomato spotted wilt virus resistance in peanut cultivar Florida-EPTM ‘113’. *BMC Genet.* 17:128. doi: 10.1186/s12863-016-0435-9
- Wang, H., Pandey, M. K., Qiao, L. X., Qin, H. D., Culbreath, A. K., He, G. H., et al. (2013). Genetic mapping and QTL analysis for disease resistance using F₂ and F₅ generation-based genetic maps derived from Tifrunner × GT-C20 in peanut (*Arachis hypogaea* L.). *Plant Genome* 6, 1–10. doi: 10.3835/plantgenome2013.05.0018
- Zhang, P., Zhu, Y. Q., Wang, L. L., Chen, L. P., and Zhou, S. J. (2015). Mining candidate genes associated with powdery mildew resistance in cucumber via super-BSA by specific length amplified fragment (SLAF) sequencing. *BMC Genomics* 16:1058. doi: 10.1186/s12864-015-2041-z
- Zhang, S., Liu, C., Wang, J. J., Ren, Z. H., Staiger, C. J., and Ren, H. Y. (2016). A processive Arabidopsis formin modulates actin filament dynamics in association with profilin. *Mol. Plant* 9, 900–910. doi: 10.1016/j.molp.2016.03.006
- Zhou, X. J., Xia, Y. L., Liao, J. H., Liu, K. D., Li, Q., Dong, Y., et al. (2016). Quantitative trait locus analysis of late leaf spot resistance and plant-type-related traits in cultivated peanut (*Arachis hypogaea* L.) under multi-environments. *PLoS ONE* 11:e0166873. doi: 10.1371/journal.pone.0166873

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Zhang, Zhang, He, Wang, Ma and Yin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.