BSA-Seq and Genetic Mapping Reveals *AhRt2* as a Candidate Gene Responsible for Red Testa of Peanut

Kun Zhang
Shandong Academy of Agricultural Sciences
Mei Yuan
Shandong Academy of Agricultural Sciences
Han Xia
Shandong Academy of Agricultural Sciences
Liangqiong He
Guangxi Academy of Agricultural Science
Jing Ma
Shandong Academy of Agricultural Sciences
Mingxiao Wang
Shandong Academy of Agricultural Sciences
Huiling Zhao
Shandong Academy of Agricultural Sciences
Lei Hou
Shandong Academy of Agricultural Sciences
Shuzhen Zhao
Shandong Academy of Agricultural Sciences
Pengcheng Li
Shandong Academy of Agricultural Sciences
Ruizheng Tian
Shandong Academy of Agricultural Sciences
Jiaowen Pan
Shandong Academy of Agricultural Sciences
Guanghui Li
Shandong Academy of Agricultural Sciences
Mahendar Thudi
Shandong Academy of Agricultural Sciences
Changle Ma
Shandong Normal University
Xingjun Wang

Shandong Academy of Agricultural Sciences
Chuanzhi Zhao (chuanzhiz@126.com)
Shandong Academy of Agricultural Sciences https://orcid.org/0000-0001-7465-7425

Research Article

Keywords: Peanut, Red testa, BSA-Seq, AhRt2, Diagnostic marker

Posted Date: December 6th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1109892/v1

Version of Record: A version of this preprint was published at Theoretical and Applied Genetics on February 15th, 2022. See the published version at https://doi.org/10.1007/s00122-022-04051-w.

Abstract

Testa color is an important trait of peanut (*Arachis hypogaea* L.). Peanuts with red testa are rich in anthocyanin, are very popular with consumers. However, genes responsible for the red testa trait in peanut are rarely reported. In order to fine map red testa gene, two F_4 populations were constructed through the cross of YZ9102 (pink testa) with ZH12 (red testa) and Zhanhong2 (red testa). Genetic analysis indicated that red testa was controlled by a single recessive gene, and named as *AhRt2* (Red testa gene 2). Using BSA-seq approach, *AhRt2* was preliminary identified in chromosome 12, and further mapped to a 530-kb interval using 220 recombinant lines through linkage mapping. Functional annotation, expression profiling, and sequence variation analyses confirmed that the anthocyanin reductase (ANR), *Arahy.IK60LM*, was the most likely candidate gene for *AhRt2.* A SNP in the third exon of *AhRt2* changed the encoding amino acids, was associated with red testa of peanut. In addition, a closely linkaged molecular marker to red testa trait was developed. Our result provide insight into the molecular mechanism underlying peanut testa color and provide valuable diagnostic marker for marker-assisted selected (MAS) breeding in peanut.

Key Message

The candidate recessive gene *AhRt2* responsible for red testa of peanut was identified thorugh combined BSA-seq and linkage mapping approaches.

Introduction

Peanut (*Arachis hypogaea* L.) is widely grown in more than 100 countries, with a total production of approximately 48.8 million tons in 2019 (http://www.fao.org/faostat/en/#data/QC). Peanut is known as an important cash crop not only for serving as high quality cooking oil but also for a variety of snacks. Peanut is superior in nutrition because it is enriched in many nutritional ingredients, such as vitamin B1, B3, B9 and E, biotin, resveratrol, isoflavones, phytic acid, anthocyanin and procyanidins, and many of which are accumulated in the testa (seed coat) (Pandey et al., 2012, Zhao et al., 2012, Zhao et al., 2020).

The testa color is an important trait of peanut and with a lot of variations, such as white, pink, red and black (or deep purple). The majority of peanut varieties have pink testa. Anthocyanin content and composition are important factors to determine the color of testa. In higher plants, there are six kinds of anthocyanins including delphinidin, cyanidin, pelargonidin, peonidin, petunidin and malvidin. Significant differences were found for the types and content of anthocyanidin in different testa color peanuts, and the content of delphinidin, cyanidin and pelargonidin were closely related to the testa color (Li et al., 2017). Recently, the metabolome comparison results suggested that the accumulation of petunidin and cyanidin was higher in red testa than that in the pink testa of peanut (Xue et al., 2021). Anthocyanins have strong antioxidant capacity and important nutritional value (Shin et al., 2006, Winkel-Shirley, 2001). High-anthocyanin varieties have become one of the important directions of rice and wheat breeding (Giordano et al., 2017, Ito & Lacerda, 2019). The testa was developed from the integument and with the

same genotype as its maternal plant, thus, the phenotype of testa color appears one generation late than other traits (Chen et al., 2021, Zhao et al., 2020). For traditional breeding methods, the phenotype of the testa color can only be identified after harvest, which prolonged the screening time. Marker-assisted selection (MAS), selecting the target progenies according to the genotype, can significantly accelerate breeding process and improve breeding efficiency, especially for testa color selection. For MAS approach, revealing the genetics and developing DNA markers tightly linked the traits are required.

With the rapid development of next-generation sequencing (NGS), considerable progress has been made in peanut whole genome sequencing including both wild type and cultivated peanuts (Bertioli et al., 2016, Bertioli et al., 2019, Chen et al., 2019, Yin et al., 2018, Zhuang et al., 2019). The availability of the completed genome sequencing of peanuts provided ideal resource for genome-wide identification of SSR and SNP markers *in silico* (Zhao et al., 2017, Ma et al., 2020). In addition, large-scale 58K SNP Array (*Axiom_Arachis*, v1) and 48K SNP Array (*Axiom_Arachis2*, v2) have been developed, which provide a new chance for genetic analysis, constructing high-resolution linkage maps, and QTL mapping (Clevenger et al., 2017, Clevenger et al., 2018b, Nabi et al., 2021). Moreover, great efforts have been made in the genetic map construction, fine mapping, and MAS of peanut (Agarwal et al., 2018, Han et al., 2018). The major gene related to the black testa color of peanut have been identified through BSA-seq and eQTL approaches (Zhao et al., 2020, Huang et al., 2020).

The previous studies suggested that peanut the red testa was controlled by one dominant gene (*R1*) and two recessive genes (*r2*, *r3*), and all these three genes appeared to inherit independently (Branch, 2011). Recently, a major dominant gene controlling red testa color, *AhRt1*, was fine-mapped on chromosome A03, and provided the closely linked makers for MAS breeding (Chen et al., 2021, Zhuang et al., 2019). However, the recessive gene in controlling the red testa color was poorly characterized.

In the present study, one recessive gene controlling red testa, *AhRt2*, was fine-mapped to a 0.5 Mb genomic region in chromosome 12 using the BSA-seq and linkage mapping approaches. An anthocyanidin reductase (ANR) gene was suggested to be the possible candidate gene. A “G/A” SNP is in the third exon of the ANR gene. We also developed tightly linked molecular markers which could be used in future MAS breeding.

Materials And Methods

Development of mapping population

For inheritance analysis and constructing the mapping populations, the pink testa peanut cultivar Yuanza 9102 (YZ9102) was used as the female parent to cross with the red testa peanut cultivars Zhonghua 12 (ZH12) and Zhanhong 2 (ZHH2), respectively (Table 1). Two populations (YZZH12 and YZZH2) were developed through single-seed-descent method. The F_1, F_2 and F_3 plants of two populations were planted in the field from 2017 to 2019 at Laixi Experimental Station （36°48'47.41" N，120°30'22.54" E）, Shandong Peanut Research Institute (SPRI), Shandong, China. The F_4 and F_5 plants were planted in the field from

2020 to 2021 at Jiyang Experimental Station of Shandong Academy of Agricultural Sciences (SAAS), Shandong, China（36°58'34.53" N，116°59'1.29" E）.

Measurement of anthocyanin

The anthocyanin content of peanut testa was measured using improved method as previous reports (Mancinelli et al., 1991, Teng et al., 2005, Zhao et al., 2020), using at least three sets of more than eight seeds per sample. In brief, frozen peanut testa (approximately 50 mg) was ground in a 5 mL centrifuge tube using liquid nitrogen. Then, homogenized testa was extracted at 4°C by adding 700 μl acidic methanol (the volume ratio of methanol to HCl is 99:1). After overnight incubation, the homogenates were centrifuged for 1 min at 12 000 rpm for 10 min. The supernatant (approximately 600 μl) was collected and mixed with 1 mL trichloromethane and 400 μl distilled water, and centrifuged at 4℃ at 12000 rpm for 10 min. Then, the absorbance of the supernatant was measured in spectrophotometer (U-3000, HITACHI, Japan) at 530 and 657 nm, respectively. The relative anthocyanin content was calculated according to the absorbance with the formula of $[A_{530} - (1/4 \times A_{657})]$ and then normalized by sample weight.

Whole genome sequencing and BSA-seq analysis

The genomic DNA was extracted from the leaves using the DNA Extraction Kit (DP305) of TIANGEN Biotech (Beijing,China) according to the manufacturer's instructions. DNA quality was determined using the BioPhotometer plus spectrophotometer (Eppendorf AG, Hamburg, Germany) and 1% agarose gel electrophoresis. For BSA-seq, two DNA pools were constructed by mixing equal amounts of DNA from 30 red testa $F_{2:4}$ individuals (Red-pool) and 30 pink testa $F_{2:4}$ individuals (Pink-pool). Then, DNA of YZ9102 and ZH12, and two DNA pools were sequenced on BGISEQ-500 platform at the Beijing Genomics Institute (BGI).

After sequencing, clean reads were obtained by removing low-quality and short reads using Soapnuke program (Chen et al., 2018), and mapped on reference genome of cultivated peanut Tifrunner (https://peanutbase.org/peanut_genome) using BWA software with the SAM tools (Li & Durbin, 2009). Single-nucleotide polymorphism (SNP) and Insertion/Deletion(InDel) were called, and filtrated by removing heterozygous and missing SNPs and InDels in the pools and parental lines using GATK software (McKenna et al., 2010). The SNP-index represents the ratio of reads harboring SNPs among the entire number of reads (Abe et al., 2012). To identify candidate regions associated with the red testa trait, the ΔSNP-index of each locus was calculated by subtracting the SNP-index of the Pink-pool from that of the Red-pool according to previous method (Takagi et al., 2013). To confirm the results of ΔSNP-index, Euclidean Distance (ED) algorithm was further preformed to identify the SNPs and InDels associated with the red testa trait using the equation reported previously (Lei et al., 2020, Hill et al., 2013).

Marker development, genetic map construction and mapping

To validate the BSA-seq results and further narrow down the region, 21 InDels in the candidate region were selected according to the comparative genomic information among the parents. Twenty one primer pairs were designed to the flanking sequences of the tragetted SNPs using Primer Premier 5.0 (http://www.premierbiosoft.com/primerdesign/). The polymorphism of these InDels was confirmed through polyacrylamide gel electrophoresis as described previously (Zhao et al., 2017). The Indel markers showing polymorphism among the parents were further used to genotype the $F_{2:4}$ population. The sequences of primers used for mapping are listed in Supplementary Table S1.

Genetic linkage map was constructed using JoinMap 5.0 software (https://www.kyazma.nl/index.php/JoinMap/). The recombinant ratio was converted into genetic distances (centimorgans, cM) through the function of Kosambi map. The linkage groups were calculated at a minimum LOD score of 5. MapChart 2.3 software was used for drawing the linkage maps (Voorrips, 2002). For QTLs analysis, inclusive composite interval mapping of additive (ICIM-ADD) was performed using software QTL IciMapping V4.1 (Meng et al., 2015). The LOD threshold was calculated by 1,000 permutations at $P < 0.05$, and the LOD score was set at 2.5 to determine the presence of a putative QTL associated with a target trait.

Prediction of candidate genes

The sequences of gene information in the candidate interval were obtained according to the cultivated peanut reference genome sequences (Version 1, https://peanutbase.org). The functions of candidate genes were annotated through Blastx program in databases of Nr (NCBI, http://www.ncbi.nlm.nih.gov), GO (https://www.geneontology.org/), KOG (http://www.ncbi.nlm.nih.gov/KOG), and KEGG (http://www.genome.jp/kegg). For sequence alignment and phylogenetic analysis of the anthocyanidin reductase (ANR) genes, the amino acid sequences of ANR proteins were download from NCBI. Multiple sequence alignments were performed by clustalw (https://www.genome.jp/tools-bin/clustalw) and BoxShade online program (https://embnet.vital-it.ch/software/BOX_form.html). The phylogenetic analysis was produced by MEGA7 (https://www.megasoftware.net/) with the neighbor-joining statistical method. The GenBank accession numbers of these proteins are provided in Supplementary Table S2.

Quantitative Real-Time PCR (qRT-PCR)

Based on Functional annotation of genes, 14 genes were selected for qRT-PCR analysis. After harvest from field, the seed coat was collected immediately before drying. The samples for qRT-PCR analysis are the same as that for BSA-seq, including two parents and RNA pools, and three biological replicates were prepared for every sample. Total RNA was extracted using Trizol Reagent kit (TaKaRa, Inc., Dalian, China) according to the instructions of manufacturer. The reverse transcriptions were performed with PrimeScript II 1st Strand cDNA Synthesis Kit (TaKaRa, Inc., Dalian, China). The primers for qRT-PCR were designed using Primer Premier 5.0 (http://www.premierbiosoft.com/primerdesign/) and listed in Supplementary Table S1. The qRT-PCR reactions were performed on ABI7500 Real Time System (USA) as

previous studies (Wang et al., 2017). The relative expressional levels of genes were calculated by $2^{-\triangle\triangle CT}$ method.

Results

Phenotyping and Genetic Analysis of Red Testa in Peanut

The phenotypic analyses showed significant differences in testa color between the two parents. For YZ9102, the testa color is traditionally pink, while ZH12 is red in testa color (Figure 1A). The red testa has more anthocyanin than that in pink one (Figure 1B). The total average content of anthocyanin in the red testa lines is about 2-7 times higher than that in the pink lines.

In order to gain insights into the genetic inheritance of red testa in peanut, the pink testa YZ9102 was crossed with the red testa Zhonghua12 (ZH12) and Zhanhong2 (ZH2) and two segregating populations, YZZH12 and YZZH2, were developed respectively (Table 1). All F_1 and F_2 seeds showed pink testa, and the F_3 seeds displayed different testa color corresponding to the coloration of either YZ9102 or ZH12/ZH2. Among the 220 F_4 individuals of YZZH12 population, 143 exhibited pink testa, and 77 showed red testa, corresponding to a segregation ratio of 5:3 by the Chi-square test ($\chi^2 = 0.59$, $p = 0.05$). Similarly, YZZH2 population, F_4 individuals also showed a 5:3 segregation ratio of pink and red ($\chi^2 = 0.69$, $p = 0.05$) (Table 1). These results demonstrated that the red testa of peanut should be controlled by a single recessive gene, and named as *AhRt2*.

BSA-seq Analysis and Mapping of Gene *AhRt2*

To map the *AhRt2* gene, we constructed two extreme pools, pink-pool and red-pool, which comprised 30 pink and 30 red $F_{2:4}$ plants, respectively. The two parents and two extreme pools were sequenced using the BGISEQ-500 platform. In total, 31.57, 41.22, 113.25 and 120.38 Gb raw data were generated for the YZ9102, ZH12, Pink-pool and Red-pool, representing approximately 12.34x, 16.12x, 47.08x and 44.29x genome coverage, respectively (Table 2). The filtered clean reads of each sample were mapped to the reference genome of the cultivar Tifrunner and a total of 412,874 SNPs/InDels were identified. To obtain the genomic region associated with the red testa, two approaches, Delta SNP and ED algorithms, were performed to calculate the allele segregation of the SNPs and InDels between the two extreme DNA pools. Delta SNP and ED algorithms showed that 54.56% and 70.11% candidate SNPs/InDels enriched on chromosome 12 (Chr.12), respectively (Figure 2A and Figure 2B). On Chr.12, the 7.8 Mb region (109.9Mb-117.7M) exhibiting significant linkage disequilibrium was identified as the candidate region for red testa (Figure 2C).

Fine Mapping of the *AhRt2* Gene

According to the BSA-seq results, 21 InDel markers were developed in the candidate region of Chr.12, and 11 of them displayed stable polymorphisms between the parental line and the F_2 individuals. These

markers were used for constructing the genetic map of candidate region and QTL mapping using the 220 $F_{2:4}$ individuals of YZZH12 population.The total length of the linkage map is 59.27 cM, which is corresponding to 110.05 Mb to 177.62 Mb region of chr.12. The major genewas mapped in a 3.64 cM region between Indel marker *InDel_16* and *InDel_20* (Figure 3). The physical location of *AhRt2* was narrowed in a 0.53 Mb region (Chr12: 117.03Mb-117.56Mb)(Figure 3).

Identification of the Candidate Genes Related to Red Testa

There are 52 genes in the candidate 0.53 Mb region of Chr.12. Among them, two genes (*Arahy.JFV18T* and *Arahy.X3FWT9*) were annotated as bHLH transcriptional factor genes, one of the core members to form the MYB-bHLH-WD40 (MBW) complex, and the latter is regarded as an important regulatory gene in plant anthocyanin biosynthesis. In addition, the candidate interval contains one anthocyanidin reductase (ANR) gene (*Arahy.IK60LM*), one of the structural genes in anthocyanin synthesis pathway, and named as *AhANR1* (Table 3). Resequencing results showed that there are 36 SNPs and 7 InDels in the candidate interval. Most of these SNPs/InDels were located in the intergenic region, and 9 SNPs were located in the gene region including two SNPs in the *AhANR1* gene region (Table 3). One SNP "*SNP_11719149*" was located in the -312 bp of the upstream of the Coding DNA Sequence (CDS). The other SNP "*SNP_117190528*" was located in the third exon of the *ANR* gene with a C/T variation, which leads to a transition from Thr (ACT) in YZ9102 to Ile (ATT) in ZH12(Figure 4A).

Sequence Alignment and Expression analysis of the Candidate Genes

AhANR1 of YZ9102had the same sequence with *A. duranensis* and *A. ipanesis*, Tifrunner, Shitouqi and Fuhuasheng, which the whole genome sequences have been released. Sequence alignment showed that threonine is conserved both in *A. duranensis* and *A. ipanesis* and other plant species, such as rice (*Oryza sativa*), common bean (*Phaseolus vulgaris*), soybean (*Glycine max)*, and tobacco (*Nicotiana tabacum*) (Figure 4B). To further analyze the candidate region, 13 genes were selected for qRT-PCR assay. Interestingly, we found that the expression of the ANR gene was up-regulated in both the red testa ZH12 and the red testa pools(Figure 4C). qRT-PCR showed that the Cт Means of the two bHLH genes (*Arahy.JFV18T* and *Arahy.X3FWT9*) exceeded to 35, indicating that the expression level of these two bHLH genes is very low in both red and pink testa. We also detected the expression level of the candidate gene of *AhRt1* in YZ9102 and ZH12, and found that the expression of *AhRt1* is very low with the Ct mean > 35. So, these two bHLH genes and the candidate gene of *AhRt1* were not likely the genes responsible for the red testa color in population used in this study. Taken together, we predicted that the *AhANR1* might be the key gene controlling the red testa.

Discussion

BSAseq is an effective strategy for gene fine mapping

BSA-seq approach is a rapid method for identifying markers linked to the traits through constructing the DNA pools with extreme traits, and first used in identifying the resistance loci in lettuce (Michelmore et al., 1991). With the development of next generation sequencing (NGS) technologies, the high throughput SNPs and InDels can be rapidly detected from the genome through sequencing. Recently, a series of NGS based BSA+ approaches have been developed, including BSA-seq (or QTL-seq), BSR-seq, MutMup and MutMup+ (Abe et al., 2012, Fekih et al., 2013, Steuernagel et al., 2016, Takagi et al., 2013). In comparison to the food crops, such as rice and wheat, the reproduction coefficient of peanut is very low. It is difficult to fine mapping the genes and QTLs through traditional mapping methods. For BSA-seq, huge segregation population is not required, and it has been successfully used in peanut for identification of the candidate interval and development linked molecular marker for many traits, such as disease resistance, shelling percentage, dormancy, black testa and red testa (Clevenger et al., 2018a, Luo et al., 2018, Pandey et al., 2017, Zhao et al., 2020, Chen et al., 2021). In this study, the candidate interval associated with the red testa was identified through the BSA-seq method. Then, the traditional mapping method was used to construct the genetic linkage map and finally narrow down to the candidate gene in a 530-kbp interval. The strategy of combined BSA-seq and traditional mapping approach were used to accelerate the identification of the candidate gene *AhRt2*, which could be used to identify other genes in peanut and other crops.

AhRt1 and AhRt2 were mapped in homologous fragment in different chromosomes

Previous studies have suggested that at least three pairs of genes were responsible for the red testa of peanut including one dominant gene and two recessive genes (Branch, 2011). Recently, *AhRt1*, a major dominant gene controlling red testa, was identified in a 580-kb interval of chromosome A03 (Chr.03). One of the bHLH transcriptional factors was suggested as the possible candidate gene for *AhRt1* (Chen et al., 2021). In the present study, a recessive gene controlling red testa color was successfully fine mapped to a physical interval of 530 kb of Chr.12, and named as *AhRt2*. Functional annotation and sequence analysis suggested that *AhRt2* could be the anthocyanidin reductase (ANR) coding gene.

Cultivated peanut is an allotetraploid with AA and BB subgenomes, and Chr.12 (B02) is corresponding to Chr.02 (A02). However, sequence alignment showed that there is no allele of *Arahy.IK60LM* in Chr.02. Instead, *Arahy.IK60LM* was mapped with the ANR gene *Arahy.W8TDEC* of Chr.03 with identity of 97.38%. Interestingly, the physical distance between *Arahy.W8TDEC* and *Arahy.07ZIFT* (*AhRt1*) is only 262 Kb. Further sequence alignment showed that most of the genes in the candidate interval (Chr12: 117.03Mb-117.56Mb) could be mapped in Chr.03, suggesting that there is a homologous chromosome fragment between Chr.03 and Chr.12. So, it is interesting that *AhRt1* and *AhRt2* were mapped in homologous fragment of different sub genome of peanut. In addition, that chromosome fragment could be found between Chr.A03 of *A. duranensis* and Chr.B02 of *A. ipaensis*, implying that the chromosome fragment exchange might have occurred before the polyploidization of cultivated peanut.

Detection of the SNPs of candidate gene in different peanut germplasm resources

To further analyze the function of the candidate gene, we detected the two SNPs of *Arahy.IK60LM* in peanut germplasm resources with different testa colors, including five pink testa, eight red testa, one white testa, two black and two black stripe peanuts. For the SNP at the upstream of the CDS, most of the germplasms are with the same genotype as the red parent, only two peanuts with black stripe with the same genotype as YZ9102. For the SNP in the third exon of the candidate gene, the “G to A” was specific only in two red parents ZH12 and ZH2 (Figure 5). It implied that there was more than one gene controlling the red testa of cultivated peanut. The molecular mechanism of the other six red testa varieties was different from that of the ZH12 and ZH2.

Development of diagnostic marker for screening the peanut with red testa

We developed a KASP marker *KASP_AhRt2* according to the sequence alignment in the candidate interval. A total of 220 lines of YZZH12 populations and 57 lines of YZZH2 populations were used for this site detection, and found that the locus of almost all red testa lines was “T:T”, and the pink testa lines were “C:C” or “T:C”, except YZZH12-127 and YZZH12-181(Figure 6). The two lines had red phenotype with “T:C” genotype. We found that the “*SNP_117190528*” of the two lines was consistent with the red parent by sequencing PCR products. Therefore, the two lines should be the linkage exchange between the *KASP_AhRt2* and *SNP_117190528.* In addition, our results suggested that the KASP marker can be used as the diagnostic genotyping marker to predict red testa peanut through MAS. However, the diagnostic marker is only used in the cross event using the ZH12 or ZH2 as donor parent for the red testa needs further validation in different genetic backgrounds.

Conclusion

In our study, one gene regulating red testa of cultivated peanut, *AhRt2*, was mapped in a 530 kb region on chromosome 12 using BSA-seq and linkage mapping approaches. Both functional annotation and sequence analysis suggested that it is a anthocyanin reductase gene. In addition, a diagnostic marker (*KASP_AhRt2*) was developed for MAS in peanut. This work lays the foundation for the further understanding of the regulation mechanisms of peanut red testa formation and molecular breeding of new varieties with red testa.

Declarations

Acknowledgements

This research is supported by National Natural Science Foundation of China (32072090, 31861143009, 31801301), Key Research and Development Project of Shandong Province (2020LZGC001), Famous subjects construction project of Shandong Agricultural and Engineering University, High-level Foreign Experts Introduction Program (GL20200123001), and Taishan Scholar Project of Shandong Province (ts20190964).

Availability of data and materials All information is specified in the manuscript or included as Additional Files.

Conflict of interest The authors declare that they have no conflict of interest.

Authors' contribution statement

Chuanzhi Zhao and Xingjun Wang conceived and designed the experiments. Mei Yuan, Jing Ma, Mingxiao Wang, Huiling Zhao and Jiaowen Pan developed the populations. Kun Zhang, Han Xia, Liangqiong He, Lei Hou, Shuzhen Zhao, Pengcheng Li, Ruizheng Tian, Guanghui Li, Mahendar Thudi and Changle Ma performed the experiments. Chuanzhi Zhao and Kun Zhang wrote the manuscript. Xingjun Wang revised the manuscript.

References

1. Abe A, Kosugi S, Yoshida K*, et al.*, 2012. Genome sequencing reveals agronomically important loci in rice using MutMap. *Nat Biotechnol* **30**, 174-8.
2. Agarwal G, Clevenger J, Pandey MK*, et al.*, 2018. High-density genetic map using whole-genome resequencing for fine mapping and candidate gene discovery for disease resistance in peanut. *Plant Biotechnol J* **16**, 1954-67.
3. Bertioli DJ, Cannon SB, Froenicke L*, et al.*, 2016. The genome sequences of Arachis duranensis and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat Genet* **48**, 438-46.
4. Bertioli DJ, Jenkins J, Clevenger J*, et al.*, 2019. The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nat Genet* **51**, 877-84.
5. Branch W, 2011. First 100 years-inheritance of testa color in peanut (*Arachis hypogaea* L.). *Crop Science* **51**, 1-4.
6. Chen H, Chen X, Xu R*, et al.*, 2021. Fine-mapping and gene candidate analysis for *AhRt1*, a major dominant locus responsible for testa color in cultivated peanut. *Theor Appl Genet* 134(11):3721-3730.
7. Chen X, Lu Q, Liu H*, et al.*, 2019. Sequencing of cultivated peanut, *Arachis hypogaea*, yields insights into genome evolution and iil improvement. *Mol Plant*.
8. Chen Y, Chen Y, Shi C*, et al.*, 2018. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* **7**, 1-6.
9. Clevenger J, Chu Y, Chavarro C*, et al.*, 2017. Genome-wide SNP genotyping resolves signatures of selection and tetrasomic recombination in peanut. *Mol Plant* **10**, 309-22.
10. Clevenger J, Chu Y, Chavarro C*, et al.*, 2018a. Mapping late leaf spot resistance in peanut (*Arachis hypogaea*) using QTL-seq reveals markers for marker-assisted selection. *Front in Plant Sci* **9**, 83.
11. Clevenger JP, Korani W, Ozias-Akins P, Jackson S, 2018b. Haplotype-based genotyping in polyploids. *Front Plant Sci* **9**, 564.

12. Fekih R, Takagi H, Tamiru M*, et al.*, 2013. MutMap+: genetic mapping and mutant identification without crossing in rice. *PLoS One* **8**, e68529.
13. Giordano D, Locatelli M, Travaglia F*, et al.*, 2017. Bioactive compound and antioxidant activity distribution in roller-milled and pearled fractions of conventional and pigmented wheat varieties. *Food Chem* **233**, 483-91.
14. Han S, Yuan M, Clevenger JP*, et al.*, 2018. A SNP-based linkage map revealed QTLs for resistance to early and late leaf spot diseases in peanut (*Arachis hypogaea* L.). *Front Plant Sci* **9**, 1012.
15. Hill JT, Demarest BL, Bisgrove BW, Gorsi B, Su YC, Yost HJ, 2013. MMAPPR: mutation mapping analysis pipeline for pooled RNA-seq. *Genome Res* **23**, 687-97.
16. Huang L, Liu X, Pandey MK*, et al.*, 2020. Genome-wide expression quantitative trait locus analysis in a recombinant inbred line population for trait dissection in peanut. *Plant Biotechnol J* **18**, 779-90.
17. Ito VC, Lacerda LG, 2019. Black rice (*Oryza sativa* L.): A review of its historical aspects, chemical composition, nutritional and functional properties, and applications and processing technologies. *Food Chem* **301**, 125304.
18. Lei L, Zheng H, Bi Y*, et al.*, 2020. Identification of a Major QTL and Candidate Gene Analysis of Salt Tolerance at the Bud Burst Stage in Rice (*Oryza sativa* L.) Using QTL-Seq and RNA-Seq. *Rice (N Y)* **13**, 55.
19. Li H, Durbin R, 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60.
20. Li H, Qiu J, Chen X, Hong Y, Liang X, 2017. Relationship between expression of anthocyanin synthesis related genes and seed coat color in peanut. *Chinese JOil Crops Sciences (Printed in chinese)* **39**, 600-5.
21. Luo H, Pandey MK, Khan AW*, et al.*, 2018. Discovery of genomic regions and candidate genes controlling shelling percentage using QTL-seq approach in cultivated peanut (*Arachis hypogaea* L.). *Plant Biotechnol J* **17(7),**1248-1260.
22. Ma J, Zhao Y, Chen H*, et al.*, 2020. Genome-wide development of polymorphic microsatellite markers and their application in peanut breeding program. *Electronic J Biotechnol* **4**, 25-32.
23. Mancinelli AL, Rossi F, Moroni A, 1991. Cryptochrome, phytochrome, and anthocyanin production. *Plant Physiol* **96**, 1079-85.
24. Mckenna A, Hanna M, Banks E*, et al.*, 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303.
25. Meng L, Li H, Zhang L, Wang JJTCJ, 2015. QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *The Crop Journal* **3**, 269-83.
26. Michelmore RW, Paran I, Kesseli R, 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *ProcNat AcadSci USA* **88**, 9828-32.

27. Nabi RBS, Cho KS, Tayade R*, et al.*, 2021. Genetic diversity analysis of Korean peanut germplasm using 48 K SNPs 'Axiom_Arachis' Array and its application for cultivar differentiation. *Sci Rep* **11**, 16630.
28. Pandey MK, Khan AW, Singh VK*, et al.*, 2017. QTL-seq approach identified genomic regions and diagnostic markers for rust and late leaf spot resistance in groundnut (i L.). *Plant Biotechnol J* **15**, 927-41.
29. Pandey MK, Monyo E, Ozias-Akins P*, et al.*, 2012. Advances in Arachis genomics for peanut improvement. *Biotechnol Adv* **30**, 639-51.
30. Shin WH, Park SJ, Kim EJ, 2006. Protective effect of anthocyanins in middle cerebral artery occlusion and reperfusion model of cerebral ischemia in rats. *Life Sci* **79**, 130-7.
31. Steuernagel B, Periyannan SK, Hernandez-Pinzon I*, et al.*, 2016. Rapid cloning of disease-resistance genes in plants using mutagenesis and sequence capture. *Nat Biotechnol* **34**, 652-5.
32. Takagi H, Abe A, Yoshida K*, et al.*, 2013. QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J* **74**, 174-83.
33. Teng S, Keurentjes J, Bentsink L, Koornneef M, Smeekens S, 2005. Sucrose-specific induction of anthocyanin biosynthesis in Arabidopsis requires the MYB75/PAP1 gene. *Plant Physiol* **139**, 1840-52.
34. Voorrips RE, 2002. MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered* **93**, 77-8.
35. Wang J, Zhang Q, Cui F*, et al.*, 2017. Genome-wide analysis of gene expression provides new insights into cold responses in *Thellungiella salsuginea*. *Frontiers in Plant Science* **8**, 713.
36. Winkel-Shirley B, 2001. Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol* **126**, 485-93.
37. Xue Q, Zhang X, Yang H*, et al.*, 2021. Transcriptome and metabolome analysis unveil anthocyanin metabolism in pink and red testa of peanut (*Arachis hypogaea* L.). *Int J Genomics* **5883901**.
38. Yin D, Ji C, Ma X*, et al.*, 2018. Genome of an allotetraploid wild peanut Arachis monticola: a de novo assembly. *Gigascience* **7**.
39. Zhao C, Qiu J, Agarwal G*, et al.*, 2017. Genome-Wide discovery of microsatellite markers from diploid progenitor species, *Arachis duranensis* and *A. ipaensis*, and their application in cultivated peanut (*A. hypogaea*). *Front Plant Sci* **8**, 1209.
40. Zhao X, Chen J, Du F, 2012. Potential use of peanut by-products in food processing: a review. *J Food Science Technology* **49**, 521-9.
41. Zhao Y, Ma J, Li M*, et al.*, 2020. Whole-genome resequencing-based QTL-seq identified AhTc1 gene encoding a R2R3-MYB transcription factor controlling peanut purple testa colour. *Plant Biotechnol J* **18**, 96-105.
42. Zhuang W, Chen H, Yang M*, et al.*, 2019. The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat Genet* **51**, 865-76.

Tables

Table 1 Genetic analysis of red testa colour in peanut

Name of the population	Generation	Total plants	Pink	Red	Expected ratio	χ^2	*P*-value
YZZH12 （YZ9102× ZhongHua 12 ）	F_4	220	143	77	5:3	0.59	0.05
YZZH2 （YZ9102× Zhanhong 2 ）	F_4	271	176	95	5:3	0.69	0.05

χ^2 3.85

Table 2 Data generated by BSA-seq

Samples	Clean reads	Clean bases	Number of reads Mapped	Mapping ratio (%)	Depth (x)	Properly paired read	Properly paired ratio
YZ9102	210,495,768	31,574,365,200	210,632,335	99.97	11.28	209,080,834	99.33%
ZH12	274,833,602	41,225,040,300	275,077,939	99.99	14.72	272,559,740	99.17%
Pink-pool	755,040,492	113,256,073,800	803,084,248	99.95	40.45	795,201,596	99.08%
Red-pool	802,571,036	120,385,655,400	754,686,636	99.83	42.99	747,421,150	98.99%

Table 3 Gene information in the candidate region on Chr.12

Gene	Start	End	Function
Arahy.1KM4NQ	117005569	117007230	MYB transcription factor MYB48
Arahy.IZM7T4	117019886	117024789	Protein kinase superfamily protein
Arahy.BL8W3W	117025116	117031218	Transcription factor jumonji (jmjC) domain-containing protein
Arahy.VDG15V	117045486	117047405	Octicosapeptide/Phox/Bem1p family protein
Arahy.B3SD72	117048799	117051828	Unknown protein
Arahy.J2UCUC	117052969	117055610	Cellulase (glycosyl hydrolase family 5)
Arahy.6S25R2	117069253	117081682	receptor-like serine/threonine kinase
Arahy.WE5WNX	117081781	117085905	receptor-like serine/threonine kinase
Arahy.KK3MU5	117089125	117096591	Unknown protein
Arahy.D43PH9	117097423	117101091	receptor-like serine/threonine kinase
Arahy.4HL1MH	117115390	117118208	Quinone reductase family protein
Arahy.X3FWT9	117131960	117135240	transcription factor bHLH123-like
Arahy.D0FPDW	117169647	117170588	transcription factor DYSFUNCTIONAL TAPETUM 1-like
Arahy.EYN48E	117173896	117177659	terpene synthase
Arahy.7M140W	117182257	117188503	FYVE zinc finger protein
Arahy.IK60LM	117188946	117191576	anthocyanidin reductase ((2S)-flavan-3-ol-forming
Arahy.ME5340	117192337	117194676	Unknown protein
Arahy.XPM13Z	117195285	117195741	Unknown protein
Arahy.6345FV	117195818	117200790	(2R)-phospho-3-sulfolactate synthase ComA
Arahy.R4UZW9	117212854	117218115	ATP binding microtubule motor family protein
Arahy.0BBN6W	117218341	117220729	Sulfite exporter TauE/SafE family protein
Arahy.92LHRM	117231502	117234533	receptor-like kinase 1
Arahy.77GZB8	117272366	117277939	branched-chain amino acid transaminase 2
Arahy.UIBC0V	117280589	117283465	branched-chain amino acid transaminase 2
Arahy.BHC7EV	117284545	117287751	Unknown protein
Arahy.FCW1MU	117311541	117316007	CBS domain-containing protein with a domain of unknown function (DUF21)
Arahy.K6XKIL	117316426	117318345	Unknown protein
Arahy.FI0236	117318361	117320009	Unknown protein
Arahy.VXEK44	117322449	117330848	LL-diaminopimelate aminotransferase
Arahy.3DBG0K	117343268	117346310	malate dehydrogenase
Arahy.JFV18T	117345716	117349561	basic helix-loop-helix (bHLH) DNA-binding superfamily protein
Arahy.WHNG6D	117365459	117375941	DHHC-type zinc finger family protein
Arahy.UW1SNI	117376910	117380446	inorganic pyrophosphatase
Arahy.T1IW9N	117377016	117379835	Unknown protein
Arahy.WTJ0L8	117380766	117380972	Unknown protein
Arahy.JM8NV3	117382006	117386697	Major facilitator superfamily protein
Arahy.7CPH5R	117414427	117417456	Glutamyl-tRNA reductase family protein
Arahy.Z3ANCQ	117417878	117423386	Unknown protein
Arahy.52ZQRU	117419135	117420277	LisH/CRA/RING-U-box domains-containing protein
Arahy.5N4DPF	117423355	117426869	Unknown protein
Arahy.TBP304	117428488	117430556	ubiquitin carboxyl-terminal hydrolase-like protein
Arahy.LX50IY	117442385	117450867	LRR and NB-ARC domain disease resistance protein
Arahy.YEAJ9J	117460320	117470837	LRR and NB-ARC domain disease resistance protein
Arahy.R7V70E	117482548	117487344	LRR and NB-ARC domain disease resistance protein
Arahy.FTDU9L	117483618	117494316	Unknown protein
Arahy.U1EPQF	117490366	117494234	LRR and NB-ARC domain disease resistance protein
Arahy.Y1XBLJ	117495399	117498854	Unknown protein
Arahy.1L886L	117501455	117505674	disease resistance protein (TIR-NBS-LRR class)
Arahy.P56WT2	117510793	117512637	Unknown protein

Arahy.FMZS4T	117543645	117547104	LRR and NB-ARC domain disease resistance protein
Arahy.4M9WLD	117555077	117560630	LRR and NB-ARC domain disease resistance protein

Table 4 SNPs in putative candidate genes in the mapping genomic region for testa color on Chr.12

SNP Position	ΔSNP	ED	Ref	Alt	Al	Genes	Gene annotation
117049317	0.94	1.33	C	G	Exon	*Arahy.B3SD72*	Uncharacterized protein
117191491	-	1.28	G	A	5'UTR	*Arahy.IK60LM*	Anthocyanidin reductase
117190528	0.95	1.35	G	A	Exon	*Arahy.IK60LM*	Anthocyanidin reductase
117328756	0.89	1.27	A	G	3'UTR	*Arahy.VXEK44*	Diaminopimelate aminotransferase
117420658	0.92	1.29	A	G	3'UTR	*Arahy.Z3ANCQ*	Protein trichome birefringence-like
117444731	0.9	1.27	A	G	Intron	*Arahy.LX50IY*	NBS-LRR disease resistance protein
117444733	0.9	1.27	A	C	Intron	*Arahy.LX50IY*	NBS-LRR disease resistance protein
117504480	0.92	1.30	T	C	Exon	*Arahy.1L886L*	NBS-LRR disease resistance protein

"Ref" is the SNP genotype in the reference genome, while the "Alt" is the genotype in the non-reference genome. "Al" indicates the functional location of identified SNP in annotated candidate gene.

Figures

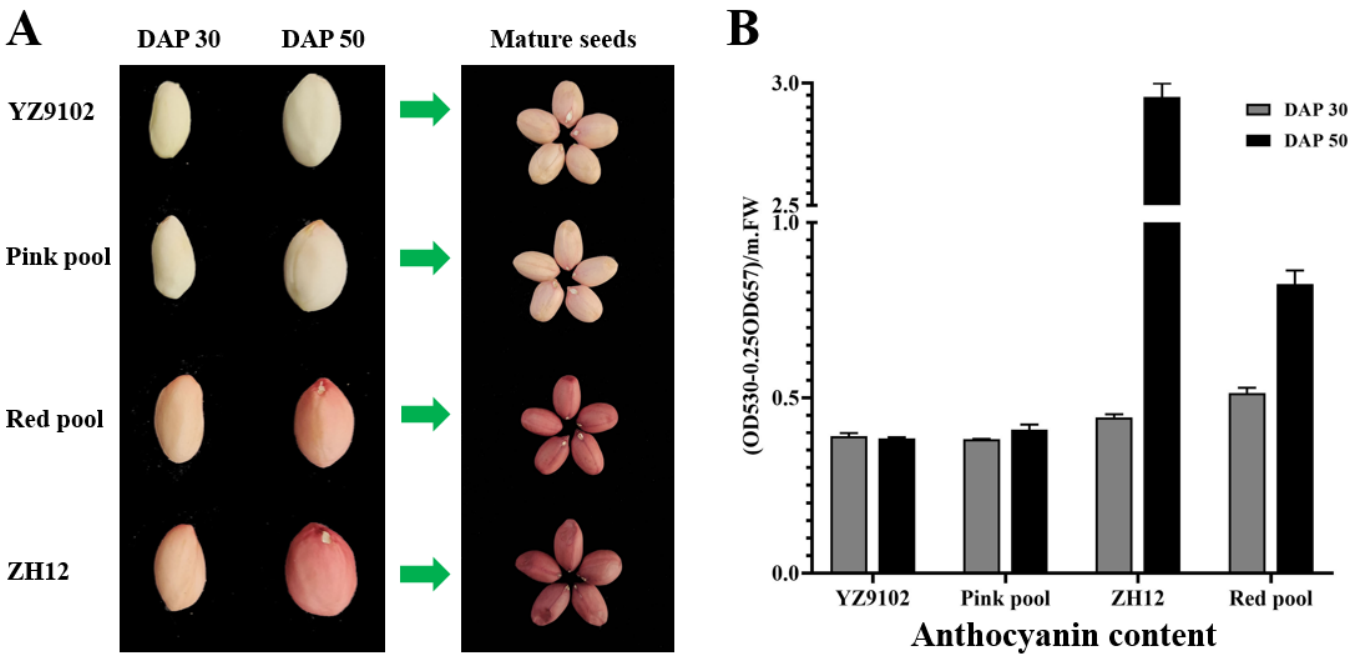

Figure 1

Phenotypes and anthocyanin content (“ DAP 30 ” means 30 days after pegging; “ DAP 50 ” means 50 days after pegging). a Seeds of pink parental line (YZ9102), red parental line (ZH12) and F4 lines with pink and red testa in different development stages, b relative anthocyanins content in parental lines and homozygous F4 lines with testa color character.

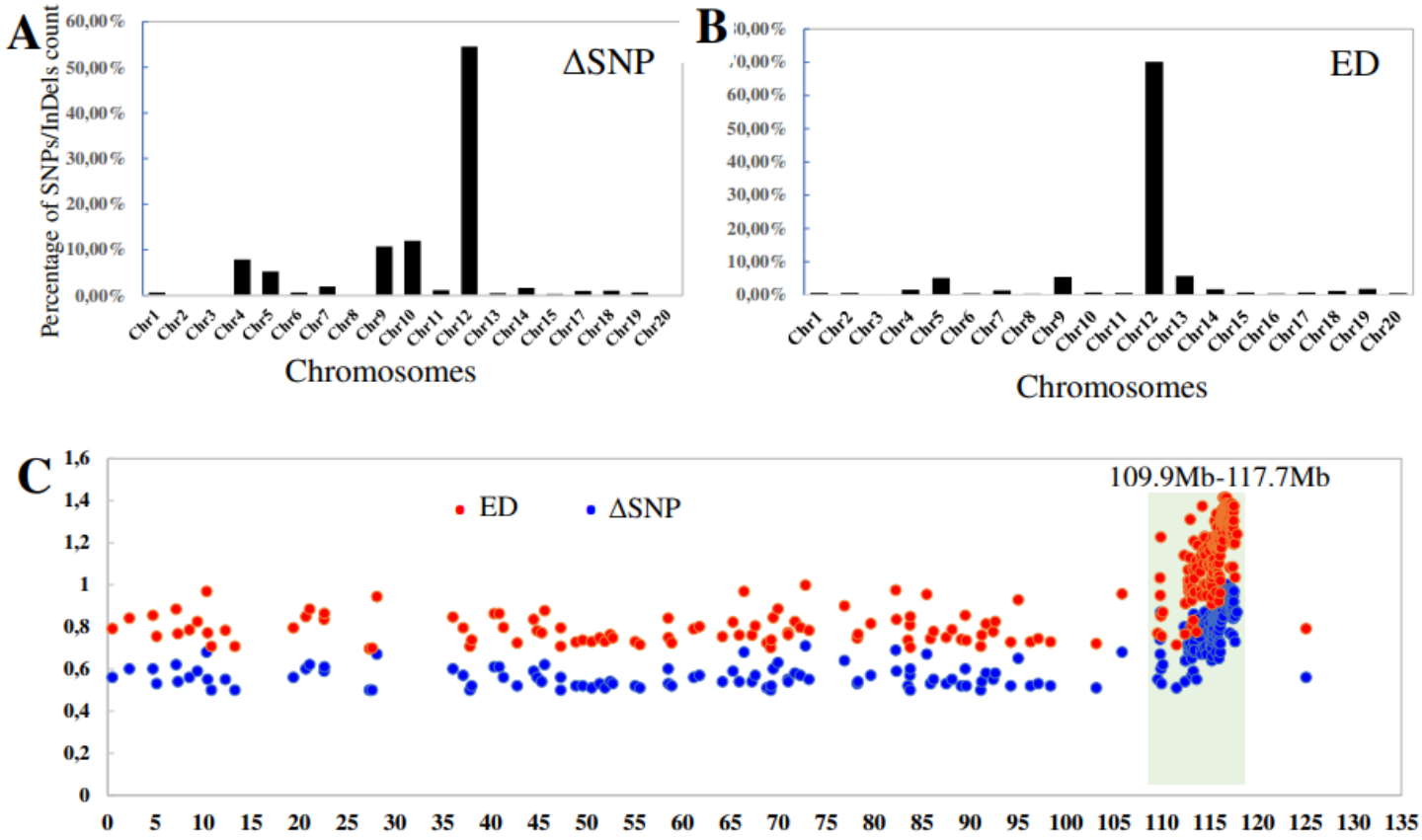

Figure 2

Distribution of candidate SNPs and InDels per chromosome. a Candidate SNPs and InDels using ΔSNP algorithm with a cutoff of ΔSNP > 0.5. b The top 1% SNPs and InDels using ED algorithm. c Distribution of candidate SNPs/InDels on Chr.12. The significant region identified for red testa pheotype is shaded (109.9–117.7 Mb).

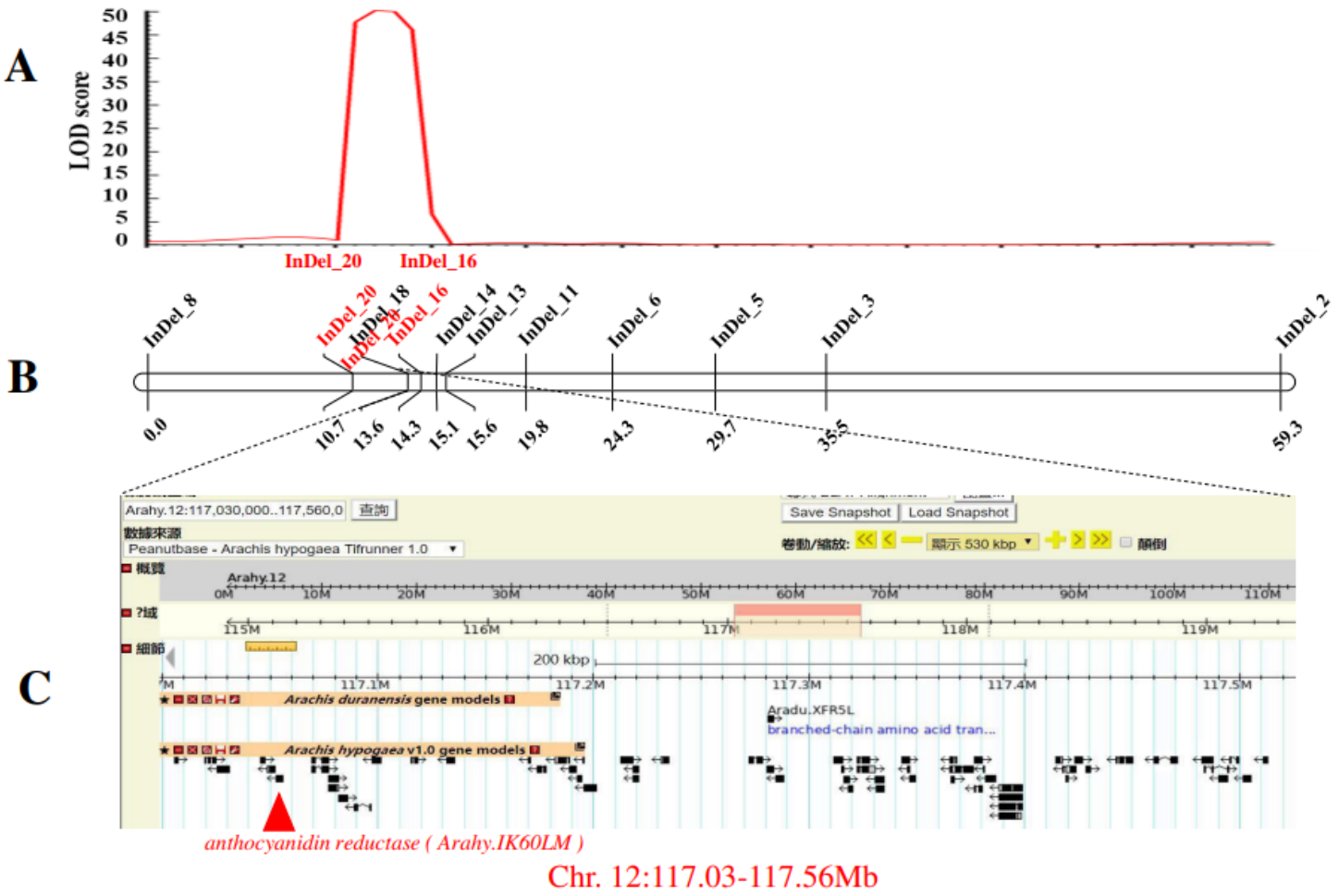

Figure 3

Identification of the candidate genes of AhRt2 through QTL mapping. a Narrowing the candidate region through IciMapping. The x-axis means the linkage groups corresponding to the candidate region of Chr.12 of peanut. The y-axis means the LOD scores. b The genetic linkage map of candidate region of Chr.12. c the genes in the 0.5 Mb interval.

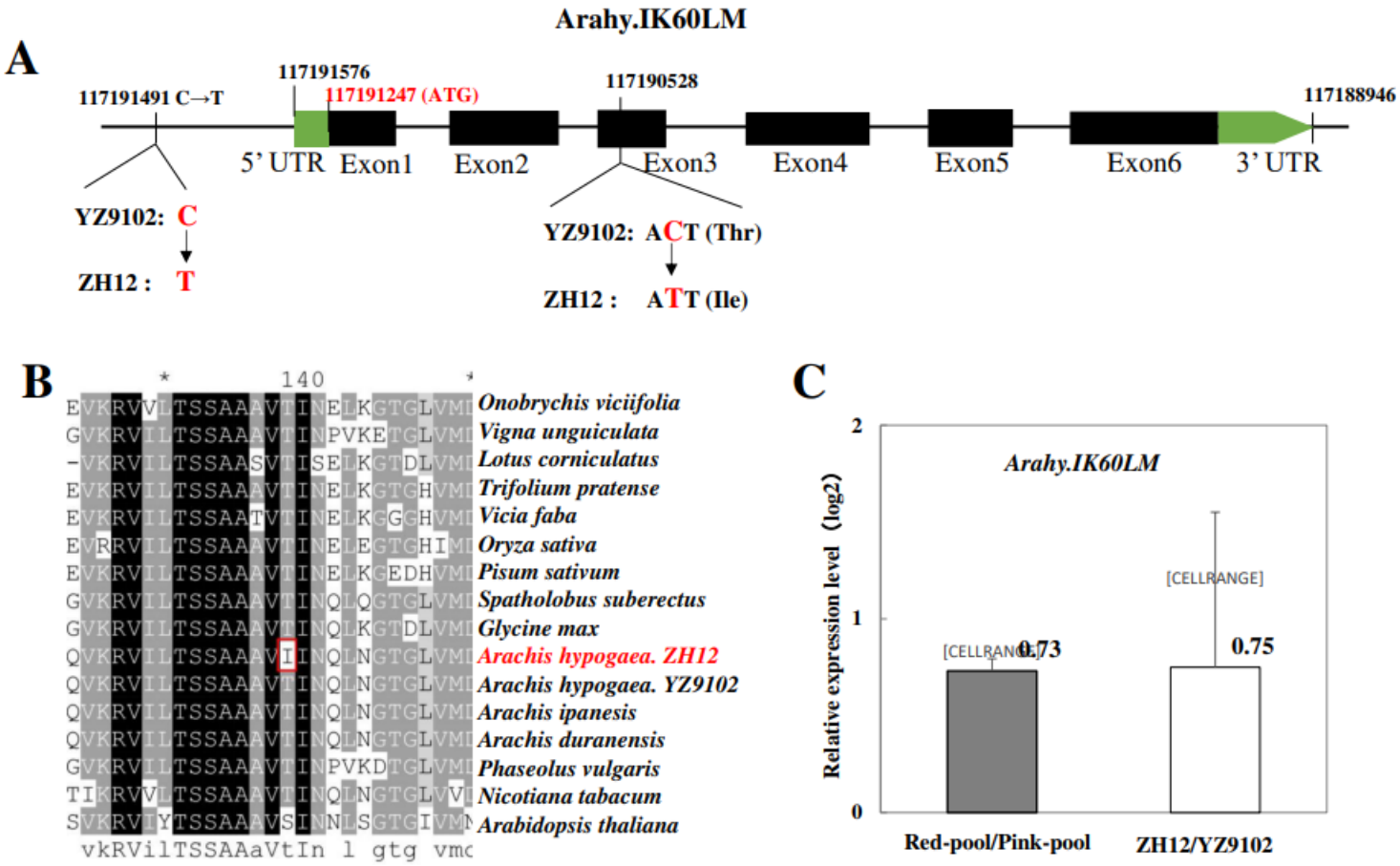

Figure 4

Gene structure and expression analysis of candidate gene. a Gene structure of candidate gene and the locations of the SNPs. b Sequence alignment of the ANR in different species. c Expression of the candidate gene in the parental lines and pools with the red and pink testa, respectively

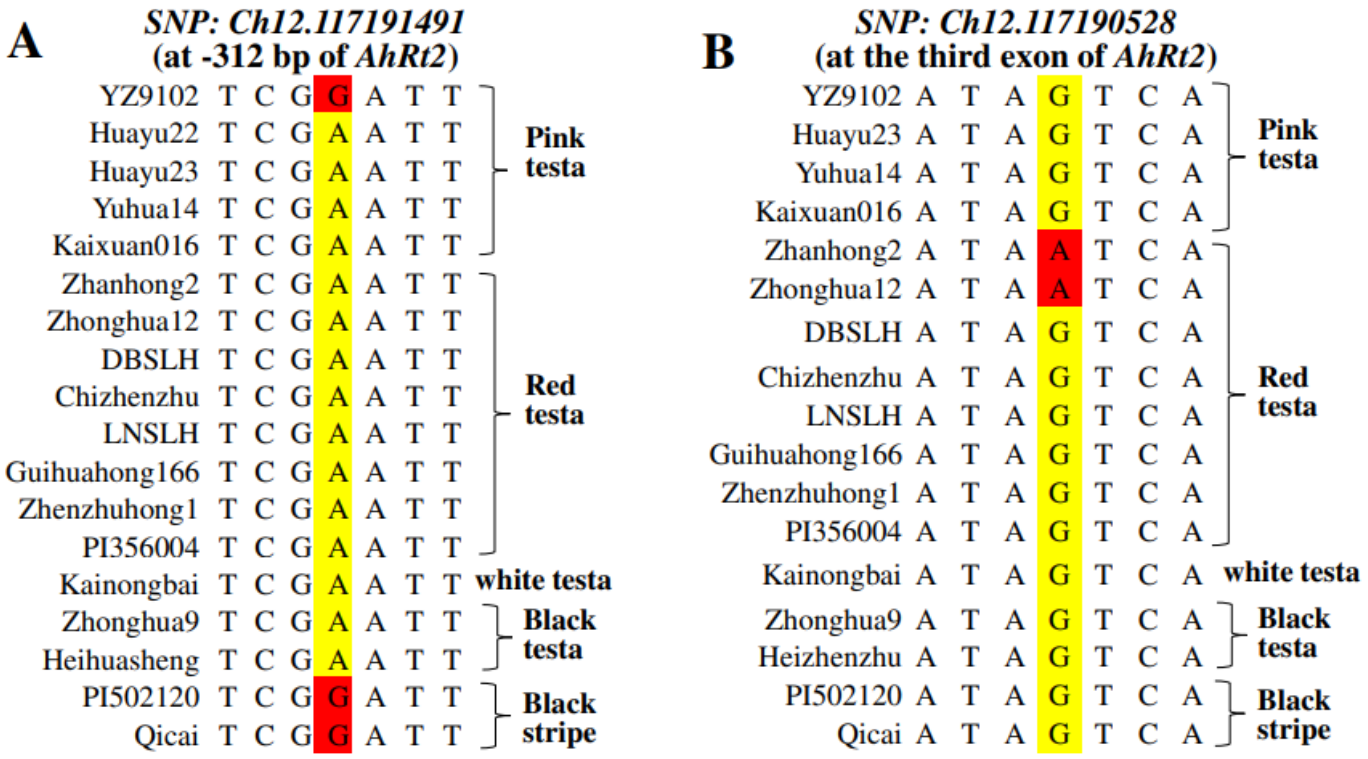

Figure 5

Detecting of SNPs of candidate gene of AhRt2 in different peanut germplasms. a Detecting of SNPs at the -312 bp of the initiation codon of the candidate gene. b Detecting of SNPs at the third exon of AhRt2.

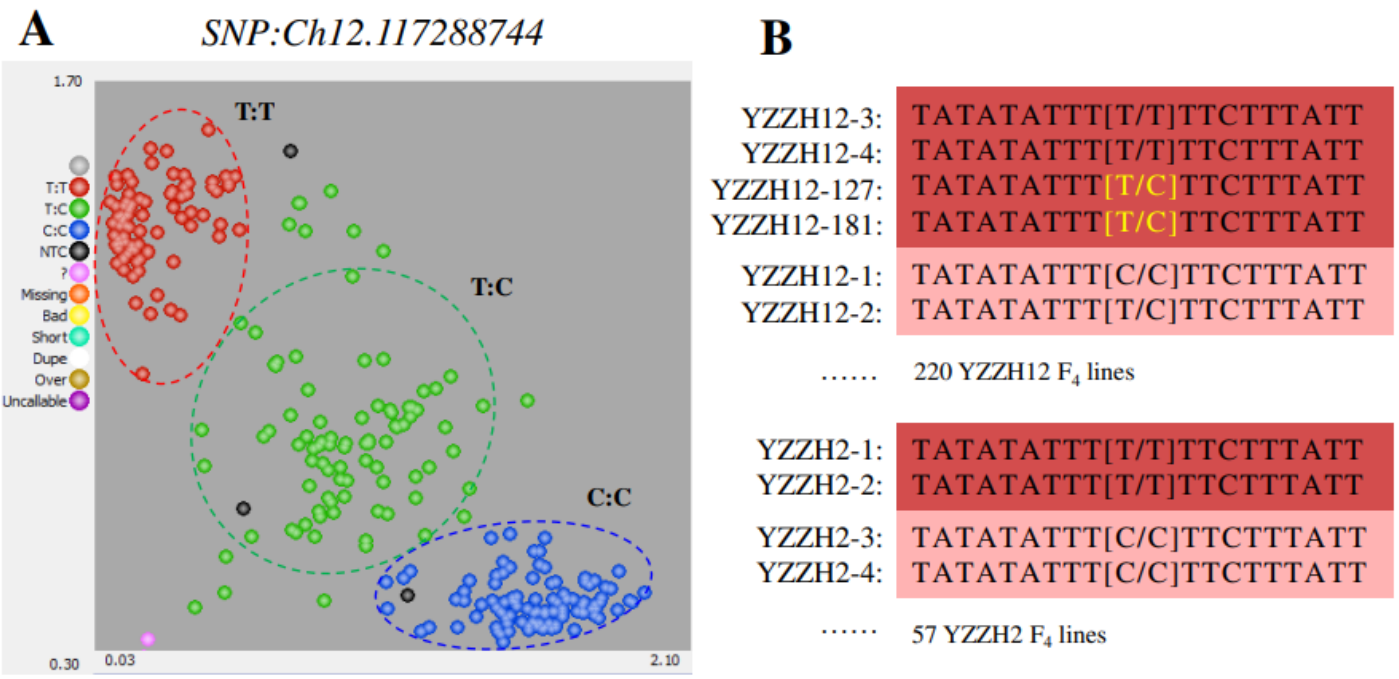

Figure 6

Genotyping results of SNP:Ch12.117288744 by KASP(Kompetitive Allele Specific PCR).a The scatter plot with axes x and y represents allelic discrimination of this site genotypes. The red, green and blue dots represent the mutant homozygous, heterozygous and wild-type homozygous, respectively. b Validation of diagnostic marker in YZZH12 and YZZH2 populations.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1Markersandprimersusedinthisstudy.xlsx