Genomics 113 (2021) 1579–1588

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Original Article

De novo full length transcriptome analysis of *Arachis glabrata* provides insights into gene expression dynamics in response to biotic and abiotic stresses

Chuanzhi Zhao [a,c,*,1], Liangqiong He [b,1], Han Xia [a,c], Ximeng Zhou [a,c], Yun Geng [a], Lei Hou [a], Pengcheng Li [a,c], Guanghui Li [a], Shuzhen Zhao [a], Changle Ma [c], Ronghua Tang [b], Manish K. Pandey [d], Rajeev K. Varshney [d,e], Xingjun Wang [a,c,*]

[a] Biotechnology Research Center, Shandong Academy of Agricultural Sciences, Shandong Provincial Key Laboratory of Crop Genetic Improvement, Ecology and Physiology, Jinan 250100, People's Republic of China
[b] Cash Crop Research Institute, Guangxi Academy of Agricultural Sciences, 530007 Nanning, People's Republic of China
[c] College of Life Sciences, Shandong Normal University, Jinan 250014, People's Republic of China
[d] Center of Excellence in Genomics & Systems Biology (CEGSB), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad 502324, India
[e] State Agricultural Biotechnology Centre, Centre for Crop and Food Innovation, Food Futures Institute, Murdoch University, Murdoch, Western Australia, Australia

ARTICLE INFO

Keywords:
Arachis glabrata
Peanut
Full-length transcriptome
PacBio
RNA-seq
Stresses

ABSTRACT

The perennial ornamental peanut *Arachis glabrata* represents one of the most adaptable wild *Arachis* species. This study used PacBio combined with BGISEQ-500 RNA-seq technology to study the transcriptome and gene expression dynamics of *A. glabrata*. Of the total 109,747 unique transcripts obtained, >90,566 transcripts showed significant homology to known proteins and contained the complete coding sequence (CDS). RNA-seq revealed that 1229, 1039, 1671, 3923, 1521 and 1799 transcripts expressed specifically in the root, stem, leaf, flower, peg and pod, respectively. We also identified thousands of differentially expressed transcripts in response to drought, salt, cold and leaf spot disease. Furthermore, we identified 30 polyphenol oxidase encoding genes associated with the quality of forage, making *A. glabrata* suitable as a forage crop. Our findings presented the first transcriptome study of *A. glabrata* which will facilitate genetic and genomics studies and lays the groundwork for a deeper understanding of the *A. glabrata* genome.

1. Introduction

Peanut or groundnut (*Arachis hypogaea*) is one of the most important oil crops in the world with a total production of approximately 48.8 million tons from 29.6 million ha during 2019 (http://www.fao.org/faostat/en/#data/QC). It provides high quality cooking oil, protein, and many other nutritional elements, such as vitamins, biotin, resveratrol and isoflavones for human consumption [1,2]. Cultivated peanut is allotetraploid (AABB, 2n = 4× = 40), derived from hybridization between *A. duranensis* (AA, 2n = 2× = 20) and *A. ipaensis* (BB, 2n = 2× = 20) [3,4]. Molecular marker analysis has demonstrated that cultivated peanut has a narrow genetic base [5], and that the overuse of elite germplasm lines in breeding programs has further reduced genetic diversity among modern cultivars. For example, more than 70% of peanut cultivars in China suggest the predominant use of two germplasms, Fuhuasheng and Shitouqi, directly or indirectly [6]. In addition, limited efforts have been made to use diverse germplasm as well as wild peanut relatives for diversification of breeding material. Based on the morphology, geographical distribution and cross compatibility relationships, more than 80 species of the genus *Arachis* were identified that can be grouped into nine sections [4,7]. Several studies have emphasized the importance of wild species as key sources for resistance to biotic and abiotic stresses, providing great potential to diversify the cultivated gene pool and enable cultivar development [8–11].

Among wild peanuts, *A. glabrata* is recognized as one of the most resistant species to disease and pests, and considered an important

* Corresponding authors at: Biotechnology Research Center, Shandong Academy of Agricultural Sciences, Shandong Provincial Key Laboratory of Crop Genetic Improvement, Ecology and Physiology, Jinan 250100, People's Republic of China.
E-mail addresses: chuanzhiz@126.com (C. Zhao), xingjunw@hotmail.com (X. Wang).
[1] These authors contributed equally to this study.

https://doi.org/10.1016/j.ygeno.2021.03.030
Received 29 July 2020; Received in revised form 21 March 2021; Accepted 29 March 2021
Available online 2 April 2021

source for genetic improvement of cultivated peanut [12]. *A. glabrata* belongs to section *Rhizomatosae*, while cultivated peanut, along with *A. duranensis*, *A. ipaensis* and *A. monticola* belong to section *Arachis*. *A. glabrata* is also known as Rhizoma peanut, creeping/perennial forage peanut or ornamental peanut grass, and provides low maintenance ground cover in many parts of the world, with bright flowers during the entire summer. *A. glabrata* has been developed as commercial tropical forage cultivars Florigraze and Arbrook due to its high protein content in the leaf and stem [13–15]. Hundreds of accessions of *A. glabrata* have been collected and maintained in Brazil, Colombia, India and the United States [16]. *A. glabrata* has demonstrated resistance to most of diseases of cultivated peanut, including leaf spots, peanut mottle virus, peanut stripe virus (PSTV), peanut bud necrosis virus (PBNV) [12], root-knot nematode [17], bacterial wilt [18] and rust [19]. However, our knowledge of the genome, ploidy and molecular mechanism of disease resistance of these species is still poorly understood, inhibiting their use in improving the resistance of cultivated peanut. To date, there are only 32 nucleotide sequences of *A. glabrata* that have been deposited in the public database (http://www.ncbi.nlm.nih.gov/genbank). Minimum genomic resources such as reference genome, transcriptome assembly and genome-wide genetic markers are therefore required to conduct experiments to enhance our understanding and use of this wild species.

Traditionally, *Rhizomatosae* species were assigned to R genome based on cytological studies. However, cytogenetic evidence shows that *A. glabrata* has a different karyotype pattern than *A. burkartii*, which is a diploid species in section *Rhizomatosae*. The karyotype features of *A. glabrata* are more comparable with species of sections *Erectoides* and *Procumbentes* (E genome) [20]. The recent availability of whole genome sequences of diploid wild ancestral species including *A. duranensis* and *A. ipaensis* [21,22], tetraploid wild species *A. monticola* [23] and three cultivars Shitouqi [24], Tifrunner [25] and Fuhuasheng [26] have provided valuable information to understand the genome and evolution of the *Arachis* genus. However, all these sequenced species belong to section *Arachis* with AA, BB or AABB genome, which have rather distant phylogenetic relationships with R or E genomes. In addition, accumulated evidence indicates that *A. glabrata* might be auto-tetraploids or segmental allotetraploids with different degrees of diploidization [20,27–29].

Transcriptome analysis is a powerful way of identifying genes, revealing the dynamics of gene expression, and providing an understanding of the molecular mechanism behind complex biological phenomena. The third-generation single-molecule real-time (SMRT) sequencing platform can capture full length transcripts and avoid the assembly process [30]. SMRT sequencing has been widely used in many plant species such as *Arabidopsis pumila* [31], rice [32,33], moso bamboo [34], loquat [35], orchid (*Bletilla striata*) [36], danshen (*Salvia miltiorrhiza* Bunge) [37] and sugarcane [38] which have provide thousands of high quality full length transcripts and facilitated the identification of alternative splicing. In this study, the upgraded PacBio SMRT Sequel platform was employed to generate a full length transcriptome of *A. glabrata*. In addition, next generation RNA-seq technology was used to investigate the gene expression dynamics of *A. glabrata* in response to biotic and abiotic stresses. Our results provide a transcriptome reference sequence of *A. glabrata* which serves as a foundation for the discovery of the molecular mechanism of stress resistance in *A. glabrata.*

2. Materials and methods

2.1. Plant materials and RNA sample preparation

A. glabrata (PI 262794) has its origin in Brazil and has high resistance to many diseases including leaf spot disease, rust, peanut mottle virus, peanut stripe virus, root-knot nematode and tomato spotted wilt viruses [17,39,40]. *A. glabrata* plants were grown in the National Wild Peanut Germplasm Garden of Guangxi Academy of Agricultural Sciences, Nanning, China. For third next generation PacBio transcriptome sequencing, the tissue samples in three biological replicates were first harvested from roots, stems, leaves, flowers, pegs, and pods of *A. glabrata* and quickly frozen in liquid nitrogen followed by storing them at −80 °C. Total RNA was extracted using RNeasy Plus Mini Kit (Qiagen), according to the manufacturer's protocol (http://www.qiagen.com). The quantity and quality of total RNA were assessed using Agilent 2100 Bioanalyzer and Fragment Analyzer Automated CE System (http://www.agilent.com).

For next generation RNA-seq, the samples of roots, stems, leaves, flowers, pegs and pods were prepared as described above. To acquire natural seedlings for biotic and abiotic treatments, rhizomes containing fresh stems and leaves were first transplanted into pots, to be grown for about two months. For salt and drought treatments, the seedlings were treated with 200 mM NaCl and 20% PEG 8000 for 48 h, respectively. For cold treatment, the seedlings were incubated at 4 °C for 48 h. For control samples, the seedlings were grown at 28 °C. For treatment with leaf spot disease, the seedlings were inoculated with *Cercosporidium personatum,* according to the method described in previous studies [41] and samples were harvested at 48 h, while samples harvested from seedlings sprayed with water were used as control. For each sample, three biological replications were prepared. For each replication, nine pots of seedlings were prepared for treatment and sample collection. After sample collection, total RNA was extracted as described above.

2.2. Library preparation and high-throughput sequencing

The PacBio SMRT sequencing library was constructed using an equal mixture of RNAs from different tissues. First-strand cDNA was synthesized using the UMI base PCR cDNA Synthesis Kit (BGI). After the synthesis of the first strand, PCR amplification was performed to generate double-strand cDNA. The cDNA was then used to select the optional size to construct a combined SMRT bell library. The library was subsequently sequenced using a PacBio sequel system. For RNA-seq, libraries were constructed using the method described in previous studies [42]. Briefly, the mRNA was first enriched and cleaved into short fragments, and then used to synthesize cDNAs. Then, cDNA fragments were purified and enriched through PCR amplification to construct the cDNA library. The sequencing was conducted using BGISEQ-500 platform (BGI, China).

2.3. PacBio Iso-Seq data processing and bioinformatic analysis

To obtain a high quality reference transcriptome of *A. glabrata*, the SMRT (IsoSeq) strategy was performed. In brief, we first obtained the "isoforms" through PacBio sequel and further treatment. These "isoforms" were used to assemble the final transcripts (reference transcriptome) of *A. glabrata*. The final transcripts were used as the reference for the following analysis, including gene expression analysis, SSR prediction and function annotation, etc. Following are the details about these steps.

After sequencing by PacBio sequel, reads of insert (ROI) were identified using SMRT analysis software package following the manufacturer's instructions. Based on the presence of 5′ primers, 3′ primers and poly (A) tail, ROIs were classified into full length (FL) and non-full length (non-FL) reads. An iterative isoform-clustering (IEC) algorithm was used to obtain the consensus sequences of isoforms. Redundant isoforms were removed through CD-HIT software, and unique isoforms (transcripts) were obtained. The completeness of the transcripts was evaluated by BUSCO (Benchmarking Universal Single-Copy Orthologs), which is based on the universal single-copy ortholog gene selected from OrthoDB database (https://www.orthodb.org/).

For annotation, the transcripts were used to align with different databases, including Nt and Nr of NCBI (http://www.ncbi.nlm.nih.gov/), GO (http://geneontology.org), KOG (http://www.ncbi.nlm.nih.gov/KOG) and KEGG (http://www.genome.jp/kegg). For CDS prediction, Trans decode (https://transdecoder.github.io) was first used to

recognize the longest ORFs (Open Reading Frame); then the ORFs were further blasted with SwissProt (http://ftp.ebi.ac.uk/pub/databases/swissprot) and Pfam (http://pfam.xfam.org). To identify the transcription factors, HMMPfam software (http://hmmer.org) was used to search the plant transcription factors database (http://plntfdb.bio.uni-potsdam.de/v3.0/). To identify the disease resistance genes, transcripts were searched with Plant Resistance (R) Genes DataBase (PRGDB), in which R genes were grouped into several classes based on the presence of specific domains (www.prgdb.org/). The SSRs were identified using Perl scripts software MISA (http://pgrc.ipk-gatersleben.de/misa/), as described in previous studies [43,44]. The primers were designed according to the flanking sequences of the SSR locus using Primer 3 software (http://biotools.nubic.northwestern.edu/Primer3.html), following the parameters described in a previous report [43].

2.4. Bioinformatics analysis of RNA-Seq data

In total, 36 samples were used for gene expression analysis, including the 18 samples from different tissues (roots, stems, leaves, flowers, pegs and pods), 12 samples for abiotic treatments and 6 samples for biotic treatment. The samples for abiotic treatments include three samples with 200 mM NaCl treatments for 48 h, three samples with 20% PEG treatment for 48 h, three samples with 4 °C low temperature treatment and three control samples. The samples for biotic treatments included three samples for leaf spot disease treatment collected at 48 h and three control samples for biotic stress. After sequencing by BGISEQ-500, clean reads were obtained by eliminating reads with poly $N > 0.5\%$, reads containing adaptor sequences, and low-quality reads using SOAPnuke (version 1.5.2) [45] with the parameter: -l 15 -q 0.2 -n 0.05. This was followed by mapping all the clean reads with the full transcripts of *A. glabrata* using Bowtie2 software [46]. Gene expression level was calculated using RSEM [47] and normalized using FPKM (Fragments Per Kilobase of transcript per Million fragments mapped) method. The relative gene expression level between different samples was calculated using R package DEGseq [48]. Differentially expressed genes (DEGs) were identified following the criteria of fold change ≥ 2 and adjusted *p*-value (q-value) ≤ 0.001. To further understand the function of the DEGs, GO analysis was performed and enrichment GO terms were identified using Blast2GO by performing hypergeometric test and comparing it with the whole transcript background [42,49]. In addition, enriched KEGG pathways were identified by comparing the ratio of DEGs with the whole transcript background.

3. Results

3.1. PacBio isoform sequencing and de novo assembly of the full length transcriptome

To acquire full length transcripts for PacBio long read sequencing, we constructed a library by combining equal quantities of total RNAs from six tissues of plants including the root, stem, leaf, flower, peg and pod of *A. glabrata* (Fig. 1). PacBio sequencing generated 22.55 Gb data including 748,874 polymerase reads with an average length of 30,107 bp. SMRT analysis identified a total of 712,736 ROIs (reads of inserts) with an average length of 1855 bp, of which 393,014 (55.14%) were full length non-chimeric reads as indicated by the detection of 5′-adapter, 3′-adapter and poly (A) tails. By using ICE (Iterative Clustering for Error Correction) algorithm, a total of 179,353 high quality isoforms were obtained. After removing the redundant sequences, a total of 109,747 unique isoforms (transcripts) were obtained which were named as Agt_1 to Agt_109747. The total length of the transcripts was 185,676,867 bp with N50 of 1974 bp and 41.11% GC content (Fig. 2A). The BUSCO assessment showed that a total of 282 complete genes (93.07%) were found, suggesting that the transcriptome assembled is relatively complete (Supplementary Fig. S1).

To further test the completeness of the transcripts for encoding the protein, the TransDecoder software was used to identify the longest open reading frame, which was further blasted with the SwissProt and

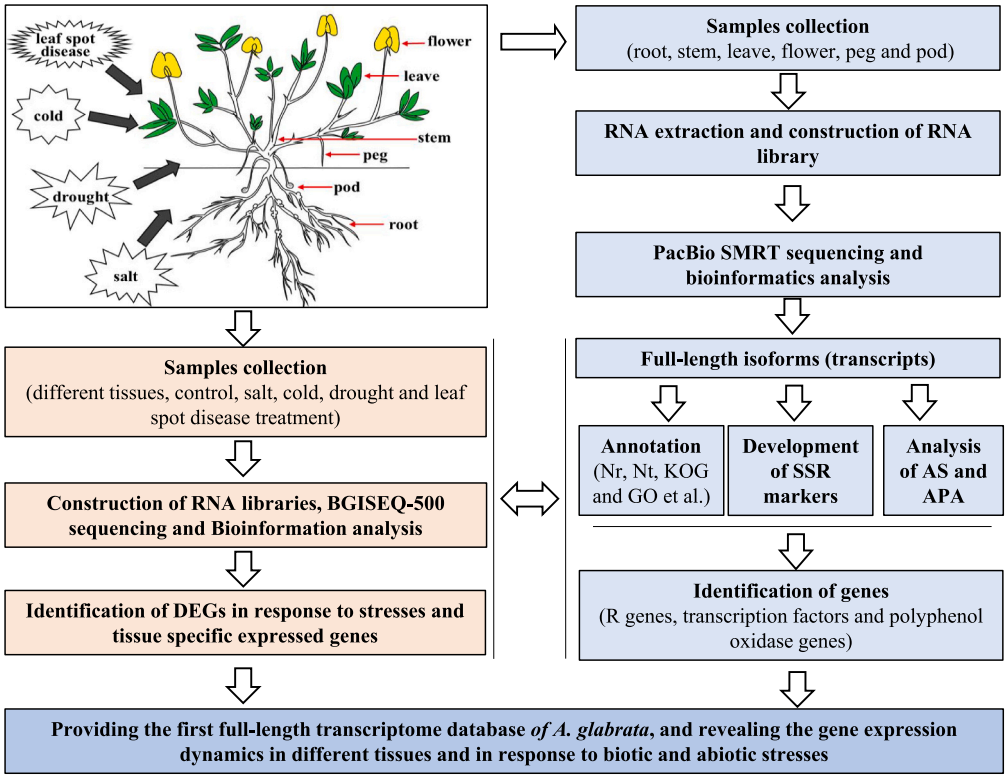

Fig. 1. Flowchart of the experimental design to obtain full length transcripts and study the dynamic gene expression profiles in different tissues and their response to various stresses.

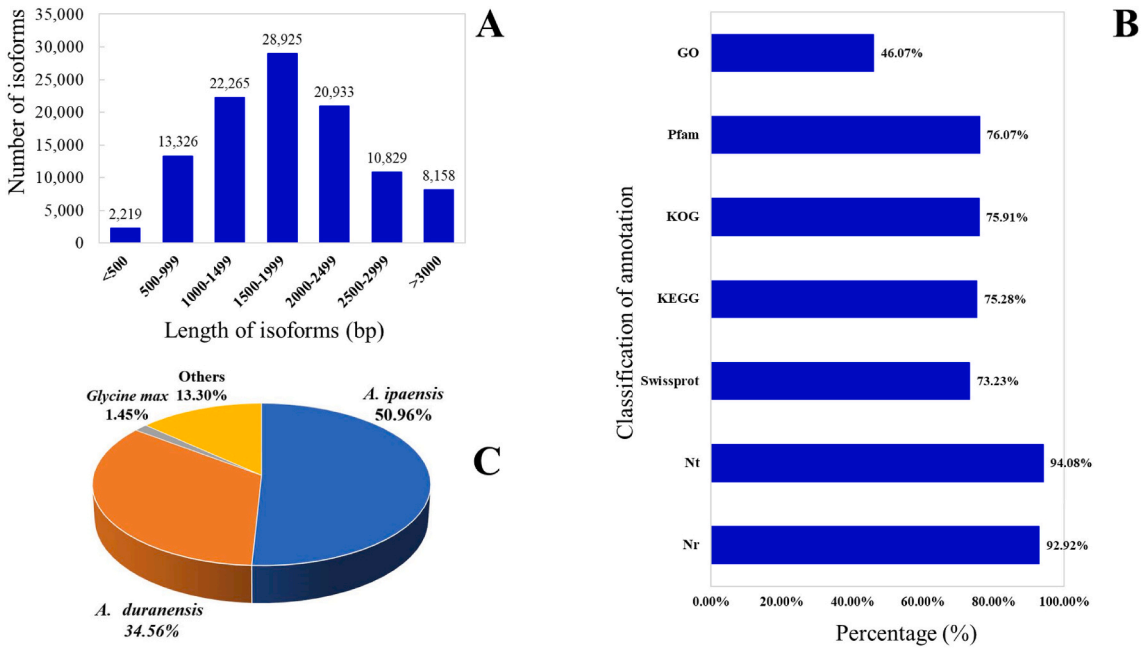

Fig. 2. Characteristics of unique isoforms (full length transcripts) of *A. glabrata.* A: Length distribution of the transcripts; B: Percentage of annotated genes in different database; and C: Species mapped with the transcripts.

Pfam Database. In total, 90,566 transcripts (85.52%) showed significant homology to known proteins and contained the complete coding sequence (CDS). The total length of CDS was more than 98 Mb with N50 of 1329 bp and 44.85% GC content.

3.2. Functional annotation of the transcripts

To acquire the most comprehensive annotation for the transcripts of *A. glabrata*, the sequences of the transcripts were used to align with Nr, Nt, KEGG, KOG and GO databases. The results showed that 101,973 (92.92%), 103,252 (94.08%), 82,619 (73.23%), 83,311 (75.91%) and 50,555 (46.07%) transcripts were potential matches for the Nr, Nt, KEGG, KOG and GO databases, respectively (Fig. 2B). Nr annotation also showed that approximately 51,963 (50.96%) and 35,242 (34.56%) transcripts showed the best matches with the gene models of the diploid ancestral species of peanut, *A. ipaensis* and *A. duranensis*, respectively. While 1478 (1.45%) transcripts matched with *Glycine* max, the rest of the 13,562 (13.30%) transcripts showed similarity with other plant species (Fig. 2C).

Transcription factors (TFs) play an important role in plant development and in responses to biotic and abiotic stresses. In *Arabidopsis*, more than 5% of the genes encoded TFs were assigned to stress related functions [50]. In this study, a total of 4276 (4.19%) transcripts encoding 59 types of TFs were characterized through blasting with plant transcription factor database (PlnTFDB). Among these TFs, MYB, AP2, GRAS, bHLH, C3H, WRKY, C2H2 and ARF accounted for 50% of the total TFs, representing the most abundant TF gene families in *A. glabrata* (Supplementary Fig. S2). The complete annotation of these transcripts (Supplementary Table S1) provides a basic framework for future studies on gene identification and gene transcriptional regulation in *A. glabrata.*

A search in the Plant Resistance Genes database (PRGdb) identified a total of 8350 disease resistance (R) related transcripts, grouped into six distinct classes based on the presence of specific domains (Supplementary Table S2). In total, 1366 and 1582 transcripts contained CC-NBS-LRR (CNL) and TIR-NBS-LRR (TNL) domains, respectively. A total of 3225 and 115 transcripts contained receptor-like protein (RLP) and receptor-like kinase (RLK) domains, respectively. In addition, 54 transcripts indicated no typical resistance related domains, but were found homologous with other reported R genes (Supplementary Table S2). *A. glabrata* was considered one of the most adaptable wild peanut species and these results provided valuable information for cloning R genes from *A. glabrata.*

3.3. Development of genomic-SSR markers

Microsatellites or simple sequence repeats (SSRs) are valuable in plant genetics and breeding studies, especially those derived from transcriptome sequences. The full length transcripts of *A. glabrata* provide important resources for SSR development. A total of 46,126 SSRs were identified from 32,934 transcripts using MISA Perl script and 9069 of these transcripts contributed to more than one SSR. The overall frequency of occurrence of SSRs was recorded as 248 SSRs per Mb with the average distance between two SSRs being 4.02 kb (Supplementary Table S3). A total of 288 types of motifs were detected in the transcripts of *A. glabrata*. Based on the flanking sequences of the identified SSRs, we designed 18,290 pairs of primers, which demonstrate their potential application value in *A. glabrata* and other peanut species (Supplementary Table S4).

3.4. Expression analyses of transcripts in various tissues of A. glabrata

To further analyze the gene expression profile in various tissues of *A. glabrata*, we used BGISEQ-500RNA-seq technology to quantify transcript abundances in the root, stem, leaf, flower, peg and pod. Approximately 63 million reads were generated from each tissue, and about 85% of them were mapped to the reference transcripts of *A. glabrata* (Table 1). The gene expression level was normalized by calculating values of fragments per kilobase of per million fragments sequenced (FPKM). We used a stringent cutoff value of FPKM ≥ 1 to define transcripts that were robustly expressed in specific tissue. Among these 109,747 transcripts, 60,689 (55.30%) in root, 62,139 (56.62%) in stem, 61,082 (55.66%) in leaf, 50,305 (45.84%) in flower, 63,687 (58.03%) in peg and 56,398 (51.39%) in pod were detected, representing the expressed genes in a specific tissue (Fig. 3). A total of 33,740 (30.74%) transcripts expressed in all six tissues collectively. In addition, a total of 1229, 1039, 1671, 3923, 1521, and 1799 transcripts were expressed specifically in the root, stem, leaf, flower, peg and pod, respectively (Fig. 3). Under the criterion of q-value ≤ 0.001 and $|\log2| \geq 1.0$, a large

number of differentially expressed genes (DEGs) were identified between different tissues (Fig. 4, Supplementary Table S5). For example, the expression level of 17,616 and 14,826 transcripts were found up- and down-regulated, respectively in the root compared to the stem (Fig. 4, Supplementary Table S5). These results indicate significant differences in the expression level of transcripts between different tissues. Detailed information on these specific transcripts provide an opportunity to study the function of the genes or cloning of tissue specific promoters.

3.5. Identification DEGs in response to abiotic stresses

In order to investigate the effects of abiotic stresses of transcript level in *A. glabrata*, RNA-seq was performed to quantify transcript abundance in response to salt, cold and drought treatments (Table 1). Compared to control samples, nearly 90,000 DEGs were identified under the criterion of q-value ≤ 0.001 and $|\log2| \geq 1.0$. We found that the total number of up-regulated DEGs was greater than down-regulated DEGs in all the three treatments investigated (Fig. 5A). For example, a total of 15,208, 12,467, and 27,426 transcripts were up-regulated in response to salt, drought and cold, respectively, while the number of down-regulated DEGs were 8117, 5857, and 20,479, respectively (Fig. 5A). The expression level of 7373 transcripts significantly changed in response to all three stress treatments (Fig. 5B). In addition, many shared DEGs were identified between two stresses, including 11,272 between salt and drought, 13,634 between salt and cold, and 11,472 between drought and cold (Fig. 5B).

A GO enrichment analysis of the DEGs identified the top 20 enriched GO categories associated with drought and salt response (Fig. 6), and most of them were related to cell structure and metabolism. Among these categories, "cell wall organization or biogenesis", "cell wall organization" and "external encapsulating structure organization" were detected as the most enriched terms in response to drought (Fig. 6). Interestingly, these three GO terms also represented the most enriched terms in salt treatment (Fig. 6). Some metabolic related GO-enriched terms such as "cellulose metabolic process" and "glucan metabolic process" were common in both drought and salt treatments (Fig. 6).

In case of low temperature stress, most of the enriched GO terms related to signal transduction (Fig. 7) which exhibited large differences from those in salt and drought stresses. In response to low temperature, "signaling receptor activity", "receptor activity", "molecular transducer activity", "DNA binding transcription factor activity", and "signal transducer activity" ranked the top five enriched GO terms. KEGG pathway enrichment analysis showed that the pathways related to circadian rhythm, photosynthesis-antenna proteins, and plant hormone signal transduction were enriched. It is worth mentioning that more than 1200 DEGs were identified in the pathway of plant hormone signal transduction which was inconsistent with GO results (Fig. 7). In addition, some metabolic-related pathways were also enriched in response to low temperature, such as carotenoid biosynthesis, lipid metabolism, thiamine metabolism, starch and sucrose metabolism, and sulfur metabolism (Fig. 7).

3.6. Identification of DEGs in response to leaf spot

To determine the genes in response to leaf spot disease, RNA-seq was performed and approximately 22 million reads were generated from samples before and after infection with leaf spot, and more than 90% of the reads were mapped to the reference transcriptome of *A. glabrata* (Table 1). A total of 56,640 and 53,697 transcripts were detected in control and treatment libraries, respectively. Based on the standard of q-value ≤ 0.001 and $|\log2| \geq 1.0$, a total of 23,160 DEGs were detected between control and treatment libraries, with 8807 up-regulated and 14,353 down-regulated genes (Supplementary Table S5). KEGG analysis showed that ten pathways were enriched in response to leaf spot disease, and most of them were involved in the metabolism, such as "starch and

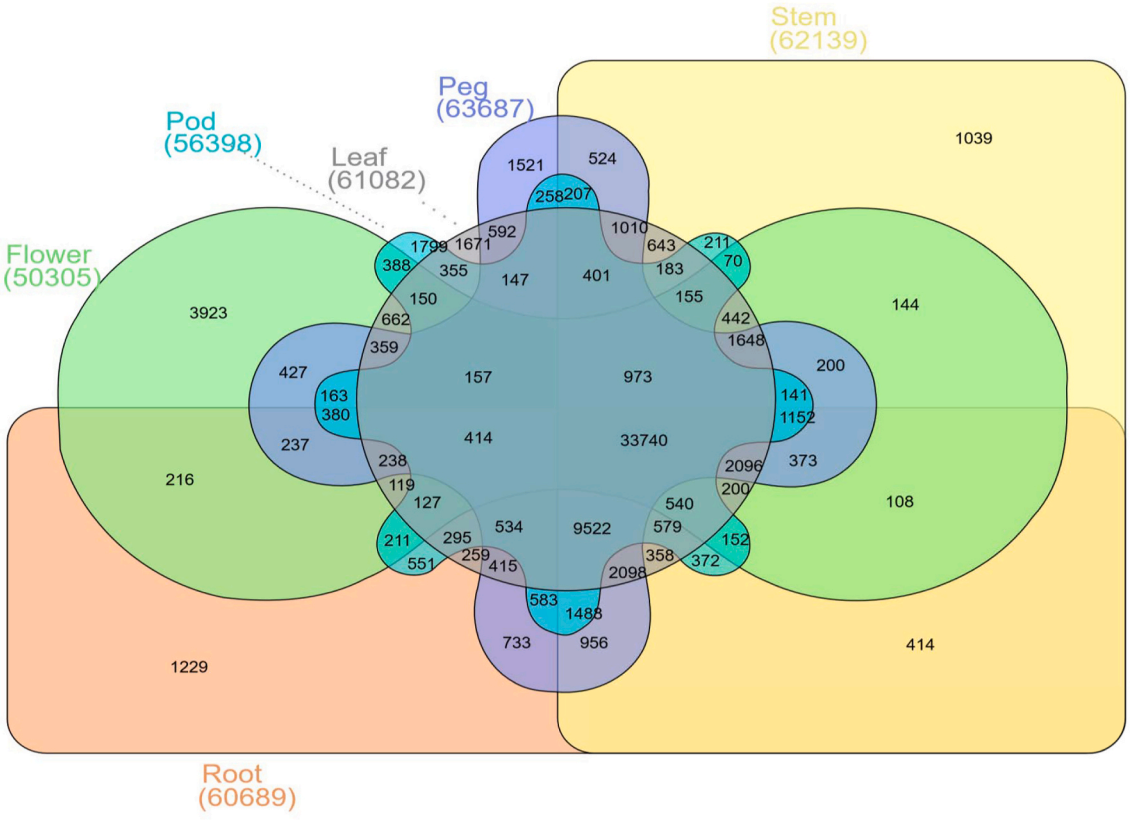

Fig. 3. Venny analysis to identify specific and common expressed transcripts in different tissues.

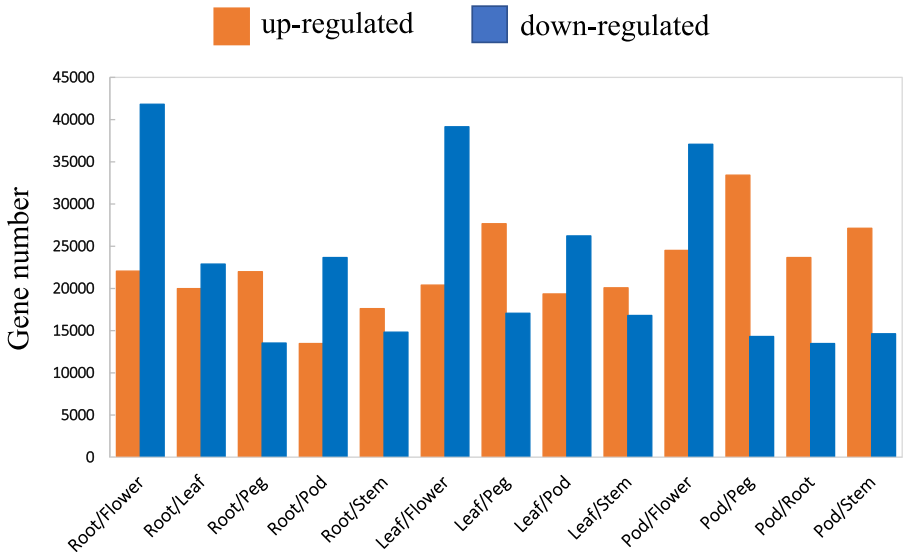

Fig. 4. Number of differentially expressed genes (DEGs) in different tissues.

Table 1
Data generated by RNA-seq.

Parts /treatments	Clean reads (Mb) and mapping in isoforms (%)		
	Replication 1	Replication 2	Replication 3
Plant parts			
Root	21.33 (89.52%)	21.21 (86.97%)	21.21 (84.12%)
Stem	21.25 (86.10%)	21.25 (91.19%)	21.30 (75.37%)
Leaf	21.20 (91.06%)	21.22 (91.65%)	21.33 (90.56%)
Flower	21.19 (88.55%)	23.02 (88.28%)	21.09 (90.77%)
Peg	21.10 (90.59%)	23.08 (89.24%)	21.09 (85.58%)
Pod	23.12 (91.68%)	21.16 (88.27%)	21.11 (88.85%)
Abiotic stress			
Abiotic stress-control (ABC)	23.02 (92.78%)	22.94 (91.92%)	21.24 (92.33%)
Cold	21.25 (86.90%)	21.18 (88.72%)	21.19 (88.37%)
Salt	21.21 (89.43%)	21.28 (90.78%)	21.28 (92.74%)
Drought	21.24 (91.48%)	21.23 (90.77%)	21.24 (90.72%)
Biotic stress (infected with leaf spot)			
Biotic stress-control (BC)	21.02 (91.77%)	22.84 (91.65%)	22.99 (92.32%)
Treatment with leaf spot (LS)	22.80 (90.54%)	23.01 (91.75%)	22.95 (91.45%)

sucrose metabolism", "amino sugar and nucleotide sugar metabolism", "glycosaminoglycan degradation", "cutin, suberine and wax biosynthesis", and "sphingolipid metabolism" (Table 2). Interestingly, many of these pathways have been reported to play important roles in plant defense responses against pathogens. For example, sphingolipids were involved in the regulation of plant programmed cell death (PCD) and defense responses [51], while cutin and wax were identified as important components perceived by invading pathogens [52–54]. In addition, "plant-pathogen interaction pathway" containing 561 DEGs was also enriched. In plants, mitogen-activated protein kinases (MAPKs) were involved in growth, development, and responses to stresses. Some MAPKs can induce the expression of defense genes and cause cell death. Here, we found that "MAPK signaling pathway" was enriched, which contained 583 DEGs (Table 2).

The NBS-LRR genes are known as one of the largest disease resistance family in plants. We analyzed the expression profiles of NBS-LRR genes in response to leaf spot disease infection. Under the standard of FPKM≥1.0 in control or treatment samples, a total 426 NBS-LRR genes were detected. Of these, 124 genes showed significant change in expression after infection with leaf spot disease under the standard of q-value ≤ 0.001 and |log2| ≥ 1.0, including 74 up-regulated and 50 down-regulated genes (Supplementary Table S6). Annotation results showed that many of these differentially expressed NBS-LRR genes were found homologous with the R genes of other species. For example, Agt_61258 encoding a TIR-NBS-LRR gene demonstrated high homology with the TMV resistance protein N transcripts X2 (XP_020987325.1) of *A. duranensis*, and was significantly induced ($\log_2 = 9$) by leaf spot disease. Similarly, Agt_2690 showed high homology with the NBS-LRR protein (MTR_5g071220) of *Medicago truncatula* and RGA1 (XP_016175578.1) of *A. duranensis*. However, the expression of Agt_2690 decreased after the plant was infected (Supplementary Table S6). *A. glabrata* is considered to be one of the most adaptable wild species and immune to leaf spot diseases. Our data provided important information to identify NBS-LRR to improve the resistance of cultivars.

4. Discussion

The *Arachis* genus contains nine sections, and so far efforts have only been made to develop genomic and transcriptomic resources for only section *Arachis*. Recognizing the value of species *A. glabrata* may contribute to the genetic improvement of cultivated peanuts as well as their use as ground cover and ornamental crops. The availability of transcriptome assembly and gene expression studies against several stresses will facilitate and encourage other such genomic and functional genomic studies. In addition, the transcriptomic resources could be crucial for testing the quality and reliability of the reference genome assembly of *A. glabrata* in future.

4.1. *De novo* full length transcriptome assembly to understand, use, and improve *A. glabrata*

The past few decades have witnessed immense progress in sequencing technology. Pacific BioSciences (PacBio) sequencing has emerged as one of the best third-generation single-molecule real-time (SMRT) platforms. Compared with second-generation technology, PacBio sequencing can generate long-read sequences and avoid the complex

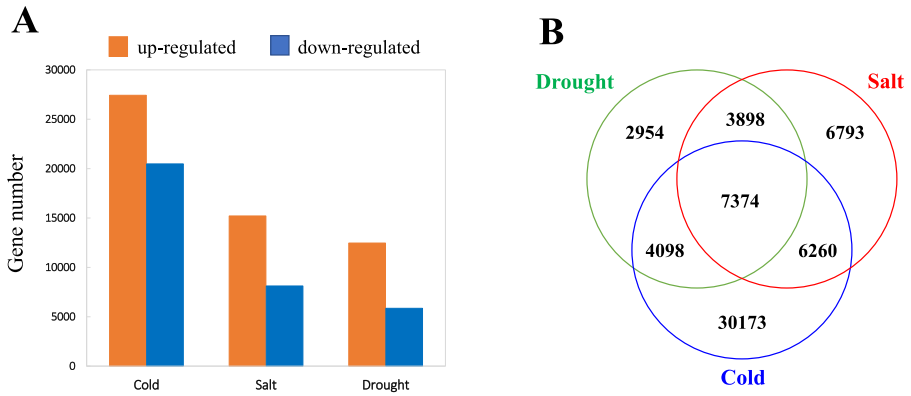

Fig. 5. Identification of DEGs in response to abiotic stresses. A: Number of DEGs in different abiotic stresses; B: Venny analysis identified common and specific DEGs in response to different stresses.

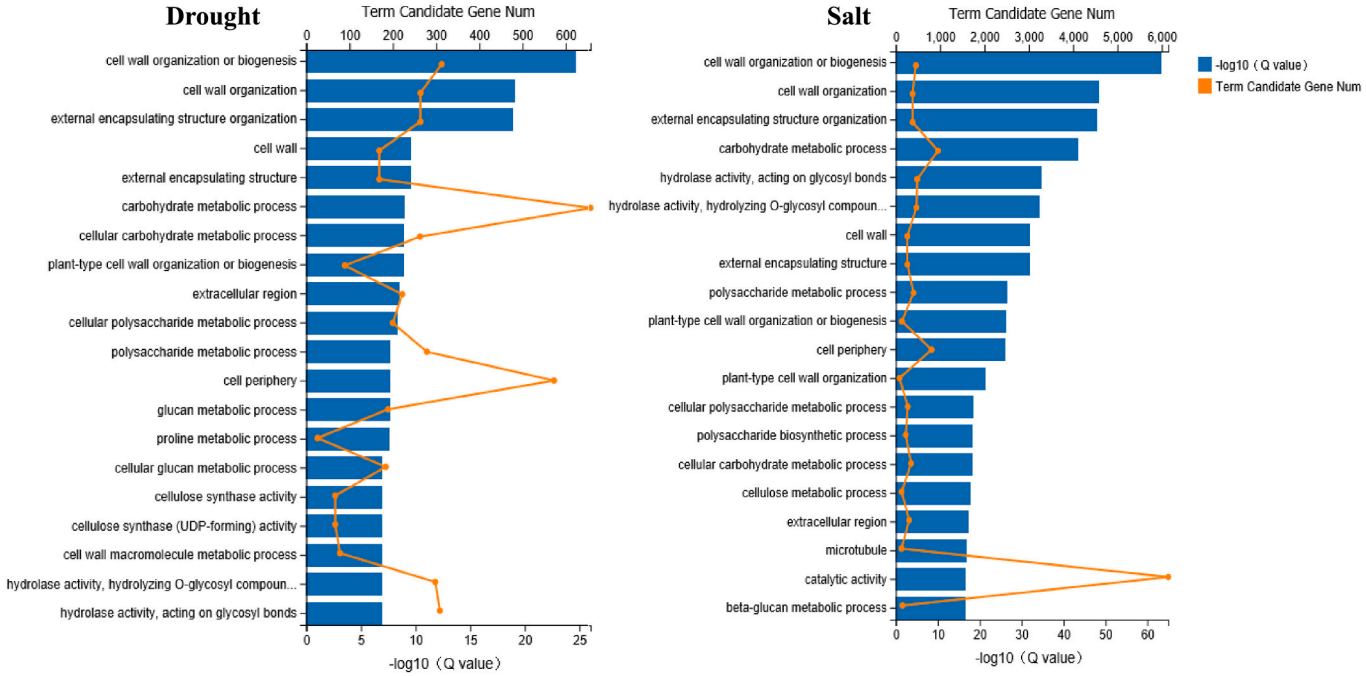

Fig. 6. GO enrichment of DEGs in response to (A) drought and (B) salt.

assembly process [30]. PacBio RS II is the first commercialized third-generation DNA sequencer providing multiple advantages in many plant species, such as maize [55], *Arabidopsis pumila* [31], moso bamboo [34] and danshen (*Salvia miltiorrhiza* Bunge) [37]. *A. glabrata* with its high resistance to biotic and abiotic stresses, is an import gene resource to improve the resistance of cultivated peanut. However, there is currently scant knowledge about genome and transcriptome. In this study, we performed the transcriptome investigation of *A. glabrata* using PacBio Sequel and obtained a reference transcriptome for it. We further analyzed the global gene expression profiles in different tissues and in response to drought, salt, low temperature, and leaf spot disease. Thousands of tissue specific genes and DEGs in response to various stresses were identified. This is the first report on the *A. glabrata* transcriptome and unveils a valuable reference for gene identification and checking the quality of the whole genome sequencing of *A. glabrata* for future research.

Although *A. glabrata*'s importance in peanut genetic improvement has long been recognized, the genetic background, ploidy, and molecular mechanisms of disease resistance remained unknown. Meiotic analysis has suggested that this species may have an autopolyploid origin with different degrees of diploidization, and its polyploid nature remains controversial [3]. In this study, we compared the transcriptome of *A. glabrata* with the genome sequences of cultivated tetraploid peanut "Tifrunner" and two diploid ancestral species of peanut, *A. ipaensis* and *A. duranensis*. Our results indicated that there was a large difference between the transcriptome of *A. glabrata* and the species of section *Arachis*. *A. glabrata* was closer to B genome than A genome.

In this study, a total of 109,747 unique transcripts of *A. glabrata* were obtained, and 90,566 transcripts were in line with known proteins. Notably, 51,963 and 35,242 transcripts consistently matched with the gene models of the diploid ancestral species of peanut, *A. ipaensis* and *A. duranensis*, respectively. The total gene numbers of *A. ipaensis* and *A. duranensis* were about 37,839 and 38,970, respectively. Based on the number of genes of the assembled transcriptome, we predicted that *A. glabrata* should be tetraploid. In another project, we determined the genome size of different *Arachis* species using flow cytometry (FCM). Our results showed that the genomes of diploid species *A. duranensis* and *A. ipaensis* were found in the range of 1.26 Gb and 1.62 Gb, respectively. The genome of *A. glabrata* is about 2.438 Gb, which is similar in size to those of the tetraploid wild species *A. monticola* (2.78 Gb) and tetraploid

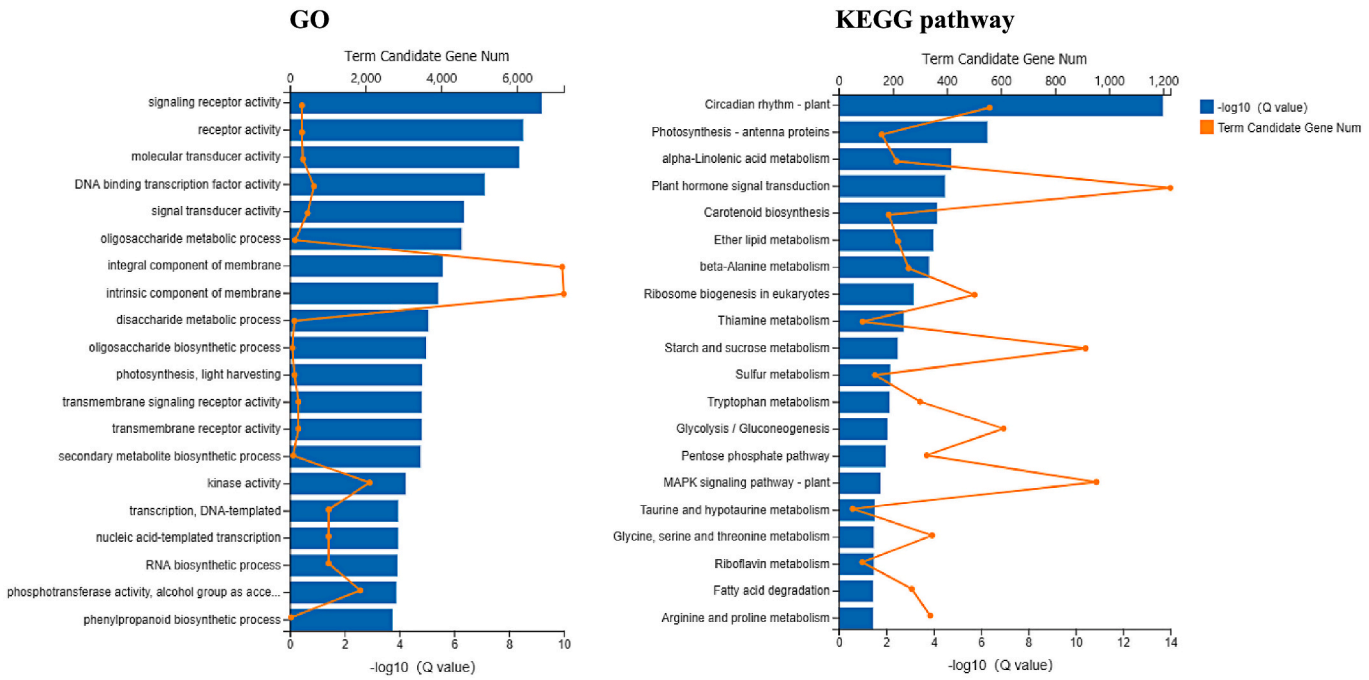

Fig. 7. GO and KEGG pathway enrichment analysis of DEGs in response to low temperature.

Table 2
Enriched KEGG pathways in response to leaf spot disease.

Pathway name	Pathway ID	Number of DEGs (%)	Number of all genes (%)	P-value
Starch and sucrose metabolism	ko00500	631 (6.12%)	1840 (4.05%)	3.34E-31
Amino sugar and nucleotide sugar metabolism	ko00520	442 (4.28%)	1384 (3.05%)	6.33E-16
MAPK signaling pathway	ko04016	583 (5.66%)	1951 (4.29%)	3.25E-14
Glycosaminoglycan degradation	ko00531	131 (1.27%)	364 (0.80%)	5.47E-09
Glycosphingolipid biosynthesis - ganglio series	ko00604	109 (1.06%)	299 (0.65%)	4.61E-08
Cutin, suberine and wax biosynthesis	ko00073	68 (0.66%)	177 (0.38%)	1.76E-06
Plant-pathogen interaction	ko04626	561 (5.45%)	2108 (0.46%)	8.22E-06
Sphingolipid metabolism	ko00600	190 (1.84%)	635 (1.39%)	1.29E-05
Ether lipid metabolism	ko00565	123 (1.19%)	384 (0.84%)	1.53E-05
Selenocompound metabolism	ko00450	79 (0.77%)	236 (0.51%)	9.58E-05

cultivated species Tifrunner (2.74 Gb). So the FCM results also suggest *A. glabrata* being tetraploid. These results will help in unraveling the mysteries of ploidy, genome, and evolution of *A. glabrata.*

4.2. Prediction of isoforms of alternative splicing (AS) and alternative polyadenylation (APA) using the reference of Arachis hypogaea L.

Alternative splicing can generate more transcripts from a single gene and therefore enhances proteome diversity [56,57]. In *Arabidopsis*, at least 42% of intron-containing genes are alternatively spliced [58]. The splicing factors are involved during most of the plant processes, particularly in the adaptation to stresses [59,60]. Alternative polyadenylation is another important way of post-transcriptional regulation for proteomic diversity and contributes to many biological processes and

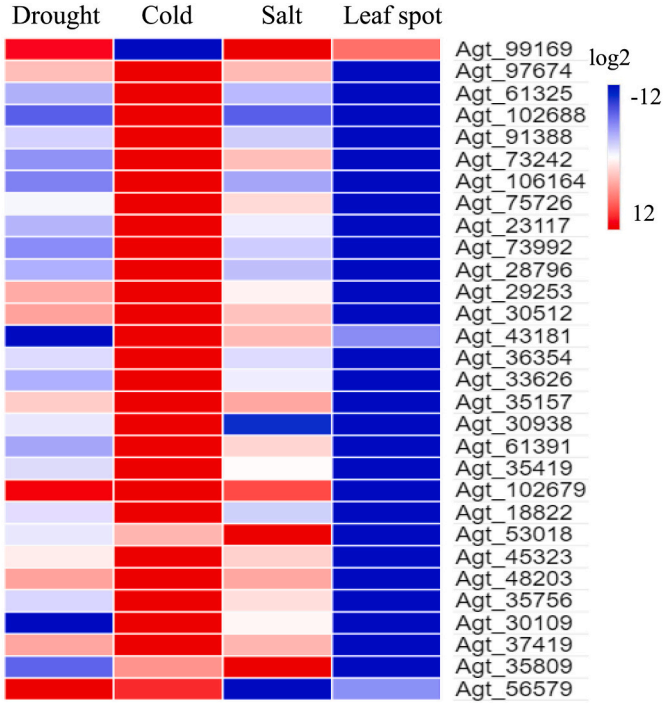

Fig. 8. Heat map showing the relative expression level of polyphenol oxidase genes in response to stresses.

stress response [61,62]. The PacBio SMRT sequencing platform can generate long reads and could capture intact mRNA sequences; hence it is very effective for the identification of AS and APA [33,34]. A total of 84,714 transcripts could be mapped with the genome of Tifrunner, and 54,847 AS events were identified by comparing with the reference genome of Tifrunner. The results showed that approximately 17,881 genes contain alternative splicing. According to the classification method described in Asprofile v1.0 (http://ccb.jhu.edu/software/ASprofile/), 12 types of AS events were identified (Supplementary

Table S7). Among them, AS of the first exon and last exon were the two most abundant types, accounting for more than 74.31% of all AS events. In addition, a total of 4690 and 996 single and multiple intron retention AS events were detected, respectively (Supplementary Table S7). A total of 35,731 polyadenylation sites from 14,384 genes were identified (Supplementary Fig. S3). Our data provides valuable resources for further studies on the variation of the proteome and the function of different spliceosome in *A. glabrata*.

4.3. Identifying polyphenol oxidases for forage quality improvement and to reveal the mechanism of resistance

A. glabrata is a good peanut species for forage production and as a cover crop for erosion control in row crops and groves, and to stabilize roadsides, engineering structures, parks, and highways. As a forage grass, its leaves contain more rumen undegraded protein (RUP) than other legume forages such as alfalfa (*Medicago sativa* L.) [63]. Greater RUP can increase the efficiency of nitrogen utilization by ruminants. Previous studies have shown that *A. glabrata* contains polyphenol oxidase (PPO) and PPO substrates that might be contributing to increasing RUP content [63]. However, the molecular mechanism governing the metabolism of PPO and RUP remained unclear. In addition, PPO has been known to be involved in plant disease resistance [64,65]. For example, overexpression of PPO in transgenic tomato enhanced bacterial disease resistance [66]. This study identified 30 PPO encoding genes, and most of them indicated resistance to biotic or abiotic stresses (Supplementary Table S8). Interestingly, most of them were down-regulated after leaf spot disease infection, while a majority of them were up-regulated in response to cold treatment (Fig. 8, Supplementary Table S8). Further functional studies of these PPO genes will provide clues to understanding *A. glabrata*'s forage quality and mechanism of disease resistance.

5. Conclusions

In this study, we used PacBio SMRT and RNA-seq technologies to study the transcriptome and gene expression dynamics of *A. glabrata*. In total, 109,747 unique isoforms were identified which represented more than 90,566 full length transcripts, thousands of tissue specific transcripts, thousands of R genes, and TFs coding genes, and 30 polyphenol oxidase genes. We also revealed the gene expression profiles of *A. glabrata* in response to drought, salt, cold, and leaf spot disease. This is the first report of *A. glabrata* transcriptome that will facilitate its genetic organization, gene cloning, and whole genome sequencing for future studies.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2021.03.030.

Funding

This research is supported by Natural Science Foundation of China (32072090, 31861143009, 31801301), High-level Foreign Experts Introduction Program of China (GL20200123001), Key Research and Development Project of Shandong Province (2020LZGC001), Guangxi Natural Science Foundation (2018GXNSFAA281079), Shandong Provincial Crop Elite Variety Development Project, Taishan Scholar Project of Shandong Province (ts20190964) and Agricultural Scientific and Technological Innovation Project of SAAS (CXGC2018E13, CXGC2016B02).

Availability of data and materials

The raw RNA sequences were available in NCBI Short Read Archive (SRA) Database (Bioproject: PRJNA674465, Accession: SRR12976768-SRR12976739). The assembled data has been deposited in NCBI Transcriptome Shotgun Assembly (TSA) Database (TSA submission: SUB8485597).

Authors contributions

XW, and CZ conceived and designed the experiments. CZ, LH, HX, XZ, YG, LH, PL, GL, and SZ performed the experiments. CZ, CM, and RT analyzed data. MKP and RKV contributed in interpretation and revising manuscript. CZ and XW drafted, revised and finalized the manuscript.

Author statement

It is our original study and not sent to anywhere else. All authors read the final version of this manuscript and agreed with the journal policy. The authors declare that they have no competing interests.

Declaration of Competing Interest

No potential conflict of interest was reported by the authors.

References

[1] M.K. Pandey, E. Monyo, P. Ozias-Akins, X. Liang, P. Guimaraes, S.N. Nigam, et al., Advances in Arachis genomics for peanut improvement, Biotechnol. Adv. 30 (2012) 639–651.
[2] X. Zhao, J. Chen, F. Du, Potential use of peanut by-products in food processing: a review, J. Food Sci. Technol. 49 (2012) 521–529.
[3] G.I. Lavia, A.M. Ortiz, G. Robledo, A. Fernandez, G. Seijo, Origin of triploid *Arachis pintoi* (Leguminosae) by autopolyploidy evidenced by FISH and meiotic behaviour, Ann. Bot. 108 (2011) 103–111.
[4] G. Kochert, H.T. Stalker, M. Gimenes, L. Galgaro, C.R. Lopes, K. Moore, RFLP and cytogenetic evidence on the origin and evolution of allotetraploid domesticated peanut, *Arachis hypogaea* (Leguminosae), Am. J. Bot. 83 (1996) 1282–1291.
[5] T. Halward, T. Stalker, E. LaRue, G. Kochert, Use of single-primer DNA amplifications in genetic studies of peanut (*Arachis hypogaea* L.), Plant Mol. Biol. 18 (1992) 315–325.
[6] B. Liao, Present situation and prospect of peanut genetic improvement in China, J. Agric. Sci. Technol. 6 (2004) 60–62.
[7] M.C. Moretzsohn, E.G. Gouvea, P.W. Inglis, S.C. Leal-Bertioli, J.F. Valls, D. J. Bertioli, 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 (2013) 113–126.
[8] S.C. Leal-Bertioli, U. Cavalcante, E.G. Gouvea, C. Ballen-Taborda, K. Shirasawa, P. M. Guimaraes, et al., Identification of QTLs for rust resistance in the peanut wild species *Arachis magna* and the development of KASP markers for marker-assisted selection, G3 (Bethesda) 5 (2015) 1403–1413.
[9] C. Simpson, Use of wild Arachis species/introgression of genes into *A. hypogaea* L, Peanut Sci. 28 (2001) 114–116.
[10] S. Pande, J.N. Rao, Resistance of wild Arachis species to late leaf spot and rust in greenhouse trials, Plant Dis. 85 (2001) 851–855.
[11] S. Sharma, M.K. Pandey, H.K. Sudini, H.D. Upadhyaya, R.K. Varshney, Harnessing genetic diversity of wild species for genetic enhancement of cultivated peanut, Crop Sci. 57 (2017) 1121–1131.
[12] N. Mallikarjuna, Gene introgression from *Arachis glabrata* into *A. hypogaea*, *A. duranensis* and *A. diogoi*, Euphytica 124 (2002) 99–105.
[13] C.M. Angelici, A.A. Hoshino, P.M. Nóbile, D.A. Palmieri, J.F.M. Valls, M. A. Gimenes, et al., Genetic diversity in section Rhizomatosae of the genus Arachis (Fabaceae) based on microsatellite markers, Genet. Mol. Biol. 31 (2008) 79–88.
[14] G. Prine, L. Dunavin, R. Glennon, R. Roush, 'Arbrook' rhizoma peanut, a perennial forage legume, in: Circular S-Florida Agricultural Experiment Stations, Institute of Food Agricultural Sciences, University of Florida, 1986.
[15] G. Prine, L. Dunavin, J. Moore, R. Roush, 'Florigraze' rhizoma peanut, a perennial forage legume, in: Circular S-Florida Agricultural Experiment Station, 1981.
[16] B.L. Maass, C. Ocampo, Isozyme polymorphism provides fingerprints for germplasm of *Arachis glabrata* Bentham, Genet. Resour. Crop. Evol. 42 (1995) 77–82.
[17] D. Baltensperger, G. Prine, R. Dunn, Root-knot nematode resistance in *Arachis glabrata*, Peanut Sci. 13 (1986) 78–80.
[18] C. Zhao, H. Xia, T. Cao, Y. Yang, S. Zhao, L. Hou, et al., Small RNA and degradome deep sequencing reveals peanut microRNA roles in response to pathogen infection, Plant Mol. Biol. Report. 33 (2015) 1013–1029.
[19] S. Oklahoma, Resistance to peanut rust in wild Arachis species, Plant Dis. 67 (1983) 209.
[20] A.M. Ortiz, G. Robledo, G. Seijo, J.F.M. Valls, G.I. Lavia, Cytogenetic evidences on the evolutionary relationships between the tetraploids of the section Rhizomatosae and related diploid species (Arachis, Leguminosae), J. Plant Res. 130 (2017) 791–807.
[21] D.J. Bertioli, S.B. Cannon, L. Froenicke, G. Huang, A.D. Farmer, E.K. Cannon, et al., The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut, Nat. Genet. 48 (2016) 438–446.

[22] X. Chen, H. Li, M.K. Pandey, Q. Yang, X. Wang, V. Garg, et al., Draft genome of the peanut A-genome progenitor (*Arachis duranensis*) provides insights into geocarpy, oil biosynthesis, and allergens, Proc. Natl. Acad. Sci. U. S. A. 113 (2016) 6785–6790.
[23] D. Yin, C. Ji, X. Ma, H. Li, W. Zhang, S. Li, et al., Genome of an allotetraploid wild peanut *Arachis monticola*: a de novo assembly, Gigascience 7 (2018).
[24] W. Zhuang, H. Chen, M. Yang, J. Wang, M.K. Pandey, C. Zhang, et al., The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication, Nat. Genet. 51 (2019) 865–876.
[25] D.J. Bertioli, J. Jenkins, J. Clevenger, O. Dudchenko, D. Gao, G. Seijo, et al., The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*, Nat. Genet. 51 (2019) 877–884.
[26] X. Chen, Q. Lu, H. Liu, J. Zhang, Y. Hong, H. Lan, et al., Sequencing of cultivated peanut, *Arachis hypogaea*, yields insights into genome evolution and oil improvement, Mol. Plant 12 (2019) 920–934.
[27] A.M. Ortiz, J.G. Seijo, A. Fernández, G.I. Lavia, Meiotic behavior and pollen viability of tetraploid *Arachis glabrata* and *A. nitida* species (section Rhizomatosae, Leguminosae): implications concerning their polyploid nature and seed set production, Plant Syst. Evol. 292 (2011) 73–83.
[28] A. Singh, C. Simpson, Biosystematics and Genetic Resources. The Groundnut Crop, Springer, 1994, pp. 96–137.
[29] M.R. Jahnavi, U.R. Murty, A preliminary pachytene analysis of two species of Arachis L, Theor. Appl. Genet. 70 (1985) 157–165.
[30] A. Rhoads, K.F. Au, PacBio sequencing and its applications, Genom. Proteom. Bioinform. 13 (2015) 278–289.
[31] L. Yang, Y. Jin, W. Huang, Q. Sun, F. Liu, X. Huang, Full-length transcriptome sequences of ephemeral plant *Arabidopsis* pumila provides insight into gene expression dynamics during continuous salt stress, BMC Genomics 19 (2018) 717.
[32] S. Schaarschmidt, A. Fischer, L.M.F. Lawas, R. Alam, E.M. Septiningsih, J. Bailey-Serres, et al., Utilizing PacBio Iso-Seq for novel transcript and gene discovery of abiotic stress responses in *Oryza sativa* L, Int. J. Mol. Sci. 21 (2020).
[33] G. Zhang, M. Sun, J. Wang, M. Lei, C. Li, D. Zhao, et al., PacBio full-length cDNA sequencing integrated with RNA-seq reads drastically improves the discovery of splicing transcripts in rice, Plant J. 97 (2019) 296–305.
[34] T. Wang, H. Wang, D. Cai, Y. Gao, H. Zhang, Y. Wang, et al., Comprehensive profiling of rhizome-associated alternative splicing and alternative polyadenylation in moso bamboo (*Phyllostachys edulis*), Plant J. 91 (2017) 684–699.
[35] C. Pan, Y. Wang, L. Tao, H. Zhang, Q. Deng, Z. Yang, et al., Single-molecule real-time sequencing of the full-length transcriptome of loquat under low-temperature stress, PLoS One 15 (2020), e0238942.
[36] L. Li, H. Liu, W. Wen, C. Huang, X. Li, S. Xiao, et al., Full transcriptome analysis of callus suspension culture system of *Bletilla striata*, Front. Genet. 11 (2020) 995.
[37] Z. Xu, R.J. Peters, J. Weirather, H. Luo, B. Liao, X. Zhang, et al., Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of Salvia miltiorrhiza and tanshinone biosynthesis, Plant J. 82 (2015) 951–961.
[38] N.V. Hoang, A. Furtado, P.J. Mason, A. Marquardt, L. Kasirajan, P. P. Thirugnanasambandam, et al., A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing, BMC Genomics 18 (2017) 395.
[39] J. Demski, G. Sowell Jr., Resistance to peanut mottle virus in Arachis spp, Peanut Sci. 8 (1981) 43–44.
[40] R. Prasada Rao, G. Ribeiro, R. Pittman, D. Reddy, J. Demski, Reaction of Arachis germplasm to peanut stripe, peanut mottle and tomato spotted wilt viruses, Peanut Sci. 20 (1993) 115–118.
[41] J. Kaur, M. Dhellon, Histological and histochemical changes in groundnut leaves inoculated with *Cercospora Arachidicola* and *Cercosporidium Personatum*, Phytoparasitica 16 (1988) 327–335.
[42] J. Wang, Q. Zhang, F. Cui, L. Hou, S. Zhao, H. Xia, et al., Genome-wide analysis of gene expression provides new insights into cold responses in *Thellungiella salsuginea*, Front. Plant Sci. 8 (2017) 713.
[43] C. Zhao, J. Qiu, G. Agarwal, J. Wang, X. Ren, H. Xia, et al., 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 (2017) 1209.
[44] R. Zhong, M. Zhou, C. Zhao, L. Hou, C. Li, X. Wang, et al., SSR marker development from peanut gynophore transcriptome sequencing, Plant Breed. 135 (2016) 111–117.
[45] P.J. Cock, C.J. Fields, N. Goto, M.L. Heuer, P.M. Rice, The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants, Nucleic Acids Res. 38 (2010) 1767–1771.
[46] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (2012) 357.
[47] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, BMC Bioinform. 12 (2011) 323.
[48] L. Wang, Z. Feng, X. Wang, X. Wang, X. Zhang, DEGseq: an R package for identifying differentially expressed genes from RNA-seq data, Bioinformatics. 26 (2010) 136–138.
[49] A. Conesa, S. Gotz, J.M. Garcia-Gomez, J. Terol, M. Talon, M. Robles, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, Bioinformatics 21 (2005) 3674–3676.
[50] J.L. Riechmann, O.J. Ratcliffe, A genomic perspective on plant transcription factors, Curr. Opin. Plant Biol. 3 (2000) 423–434.
[51] R. Berkey, D. Bendigeri, S. Xiao, Sphingolipids and plant defense/disease: the "death" connection and beyond, Front. Plant Sci. 3 (2012) 68.
[52] C. Blanc, F. Coluccia, F. L'Haridon, M. Torres, M. Ortiz-Berrocal, E. Stahl, et al., The cuticle mutant eca2 modifies plant defense responses to biotrophic and necrotrophic pathogens and herbivory insects, Mol. Plant-Microbe Interact. 31 (2018) 344–355.
[53] M. Serrano, F. Coluccia, M. Torres, F. L'Haridon, J.P. Metraux, The cuticle and plant defense to pathogens, Front. Plant Sci. 5 (2014) 274.
[54] C. Ziv, Z. Zhao, Y.G. Gao, Y. Xia, Multifunctional roles of plant cuticle during plant-pathogen interactions, Front. Plant Sci. 9 (2018) 1088.
[55] B. Wang, E. Tseng, M. Regulski, T.A. Clark, T. Hon, Y. Jiao, et al., Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing, Nat. Commun. 7 (2016) 11708.
[56] B.R. Graveley, Alternative splicing: increasing diversity in the proteomic world, Trends Genet. 17 (2001) 100–107.
[57] T.W. Nilsen, B.R. Graveley, Expansion of the eukaryotic proteome by alternative splicing, Nature 463 (2010) 457–463.
[58] S.A. Filichkin, H.D. Priest, S.A. Givan, R. Shen, D.W. Bryant, S.E. Fox, et al., Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*, Genome Res. 20 (2010) 45–58.
[59] A.B. James, N.H. Syed, S. Bordage, J. Marshall, G.A. Nimmo, G.I. Jenkins, et al., Alternative splicing mediates responses of the *Arabidopsis* circadian clock to temperature changes, Plant Cell 24 (2012) 961–981.
[60] A.M. Mastrangelo, D. Marone, G. Laido, A.M. De Leonardis, P. De Vita, Alternative splicing: enhancing ability to cope with stress via transcriptome plasticity, Plant Sci. (2012) 185–186 (40-9).
[61] D. Xing, Q.Q. Li, Alternative polyadenylation and gene expression regulation in plants, Wiley Interdiscip. Rev. RNA 2 (2011) 445–458.
[62] D.C. Di Giammartino, K. Nishida, J.L. Manley, Mechanisms and consequences of alternative polyadenylation, Mol. Cell 43 (2011) 853–866.
[63] M.L. Sullivan, J.L. Foster, Perennial peanut (*Arachis glabrata* Benth.) contains polyphenol oxidase (PPO) and PPO substrates that can reduce post-harvest proteolysis, J. Sci. Food Agric. 93 (2013) 2421–2428.
[64] P. Thipyapong, M.D. Hunt, J.C. Steffens, Antisense downregulation of polyphenol oxidase results in enhanced disease susceptibility, Planta 220 (2004) 105–117.
[65] C.P. Constabel, R. Barbehenn, Defensive Roles of Polyphenol Oxidase in Plants. Induced Plant Resistance to Herbivory, Springer, 2008, pp. 253–270.
[66] L. Li, J.C. Steffens, Overexpression of polyphenol oxidase in transgenic tomato plants results in enhanced bacterial disease resistance, Planta 215 (2002) 239–247.