

RESEARCH ARTICLE

Open Access



Global Methylome and gene expression analysis during early Peanut pod development

Pengfei Wang^{1,2}, Suhua Shi³, Junjie Ma⁴, Hui Song¹, Ye Zhang¹, Chao Gao¹, Chuanzhi Zhao^{1,3}, Shuzhen Zhao^{1,3}, Lei Hou¹, Javier Lopez-Baltazar⁵, Shoujin Fan³, Han Xia^{1,3*} and Xingjun Wang^{1,3,4*}

Abstract

Background: Early peanut pod development is an important process of peanut reproductive development. Modes of DNA methylation during early peanut pod development are still unclear, possibly because its allotetraploid genome may cause difficulty for the methylome analysis.

Results: To investigate the functions of the dynamic DNA methylation during the early development of the peanut pod, global methylome and gene expression analyses were carried out by Illumina high throughput sequencing. A novel mapping strategy of reads was developed and used for methylome and gene expression analysis. Differentially methylated genes, such as nodulin, cell number regulator-like protein, and senescence-associated genes, were identified during the early developmental stages of the peanut pod. The expression levels of gibberellin-related genes changed during this period of pod development. From the stage one (S1) gynophore to the stage two (S2) gynophore, the expression levels of two key methyltransferase genes, DRM2 and MET1, were up-regulated, which may lead to global DNA methylation changes between these two stages. The differentially methylated and expressed genes identified in the S1, S2, and stage 3 (S3) gynophore are involved in different biological processes such as stem cell fate determination, response to red, blue, and UV light, post-embryonic morphogenesis, and auxin biosynthesis. The expression levels of many genes were co-related by their DNA methylation levels. In addition, our results showed that the abundance of some 24-nucleotide siRNAs and miRNAs were positively associated with DNA methylation levels of their target loci in peanut pods.

Conclusion: A novel mapping strategy of reads was described and verified in this study. Our results suggest that the methylated modes of the S1, S2, and S3 gynophore are different. The methylation changes that were identified during early peanut pod development provide useful information for understanding the roles of epigenetic regulation in peanut pod development.

Keywords: Peanut, Gynophore, Pod development, Methylome, Small RNA

Background

Peanut is an important crop for oil and protein production in the tropic and subtropic regions worldwide. Peanut pod development is an important process of peanut reproductive development [1–3]. The peanut gynophore, the fertilized and elongated ovary, carries and eventually pushes the fertilized ovules into the soil where seed and

pod development are completed [4, 5]. Within the aerial grown gynophore, pre-embryo development is arrested until the top region of the ovary is buried in the soil [4]. In early peanut embryo development, peanut endosperm can be observed [3, 6]. Similar to other dicotyledonous species, the endosperm is absorbed during the late developmental stages [7].

Previous studies indicated that light is the major inhibitor preventing embryo and pod development before the gynophore penetrates the soil [8, 9]. Light signals can be transduced through phytohormones to

* Correspondence: coldxia@126.com; xingjunw@hotmail.com

¹Biotechnology 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
Full list of author information is available at the end of the article



regulate plant growth and development, together with gibberellin (GA), abscisic acid (ABA), brassinosteroids, and ethylene [10–12]. Previous studies have indicated that phytohormones play important roles in the early development of peanut pods. For example, auxin affected the gravitropic growth of the gynophore [1]. However, the molecular events that regulate early peanut pod development remain unknown. DNA methylation is an important epigenetic marker and a conserved mechanism for gene regulation in both plants and animals [13]. DNA methylation modulates gene expression, controls invading viruses, silences transposable elements, and maintains genomic integrity in eukaryotes [14–16]. In plants, DNA methylation occurs at three sequence contexts, CG, CHG, and CHH [13, 17–19], and is mainly determined by both DNA methyltransferases and demethylases [16, 19]. In plants, there are two types of DNA methyltransferases for the maintenance of DNA methylation: methyltransferase (MET) and chromomethylase (CMT) [20, 21]. METs maintain CG methylation, while CMTs maintain CHH and CHG methylation [22, 23]. Domains Rearranged Methyltransferase (DRM) is the key enzyme for de novo DNA methylation and requires targeting information, which is often derived from small RNA pathways [22]. Pol IV derived long dsRNAs are processed by DICER-LIKE 3 (DCL3) into siRNAs and exported to the cytoplasm. Following the loading of one strand of siRNAs onto ARGONAUTE4 (AGO4), they are imported to nucleus, where siRNAs guide the targeting by sequence complementarity. Eventually, DNA methyltransferase is guided to target loci by AGO4 associated siRNAs [24].

DNA methylation and de-methylation, including repressor of silencing 1 (ROS1) and demeter (DME), are critical for Arabidopsis embryogenesis and seed viability [25]. Specifically, ROS1 induces DNA de-methylation in extensive tissues and DME preferentially induces DNA de-methylation in central cells [26]. The apical cell of the *met* mutant embryo divides incorrectly and the viability of the seed is reduced. Many embryo-specific genes are expressed abnormally in the *met* mutant embryo [27]. DNA methylation regulates the ripening process of tomato fruit through hypermethylation of the promoter of a colorless non-ripening gene [28].

The DNA methylation patterns of the whole peanut genome and the functions of DNA methylation on peanut pod development have not been reported. In this study, we analyzed DNA methylation of the peanut gynophore/pod at different developmental stages including (1) above ground green gynophores, (2) white gynophores that have been buried in soil for approximately 3 days, and (3) gynophores with a just enlarged pod that have been buried in soil for approximately 9 days. Dynamic DNA methylome and gene expression analyses

of these three stages facilitated the investigation of the roles of DNA methylation on peanut pod development. Our results demonstrated that genes that may play key roles in embryogenesis are differentially methylated in these three developmental stages.

Methods

Plant materials

Peanut cultivar Luhua14 was used for this study. The criteria for gynophore/pod staging followed methods described in previous studies [29, 30]. Three millimeter tips of stage 1 (S1) and stage 2 (S2) gynophores and the enlarged ovary of the stage 3 (S3) gynophore were collected for DNA isolation.

Genomic DNA extraction, library construction, and high throughput sequencing

Genomic DNA was extracted following a previously described method [31]. DNA was used for MeDIP libraries construction and each MeDIP library was subjected to high throughput sequencing by an Illumina solexa HiSeq2000 platform following previously described methods [31]. Two biological replicates were used for the DNA methylation analysis.

Sequencing quality control and reads mapping

Quality control was performed using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).

Clean reads were mapped on the A and B sub-genome (pse-genomes) using Bowtie2 (version 2.0.5) software with default parameters [32]. Peanut pse-genome sequences were obtained from the peanut genome database (<http://peanutbase.org/>).

The identification of orthologs between pse-a and B genomes

First, all protein coding genes of the pse-A and B genomes were identified by local all-vs.-all BLASTP (E-value < 10^{-20}). The best-hit gene-pairs between the pse-A and B genomes were selected for subsequent analysis. Secondly, interspecies synteny analysis of the pse-A or B genomes was based on comparisons of 100 kb chromosome blocks containing the genes of the best-hit gene-pairs [33, 34]. When three or more conserved homologous gene-pairs were individually detected in the two blocks from the pse-A and B genomes these two blocks were considered syntenic blocks [33–35], and the pairs of genes were considered orthologous between the pse-A and B genomes.

Novel mapping strategy

Clean reads were mapped onto both the pse-A and B genomes. We first mapped one clean read to the pse-A genome, and then mapped the same read to the pse-B genome. Based on the mapping results, we could identify if this

read could be mapped to pse-A, or pse-B, or both genomes. The mapping results were important for the next calculation of reads of genes per million mapped reads (RPM) of a particular gene. For this, we first confirmed whether the target gene is pse-A genome specific, pse-B genome specific, or present in both the pse-A and B genomes. For the pse-A or B genome specific genes, we counted the real number of reads that mapped to this gene for the calculation of RPM.

If the gene was present in both the pse-A and B genomes, it was considered orthologous for that target gene, and all reads that mapped to the orthologs of the pse-A and B genomes were counted for the calculation of RPM. However, two types of reads existed when they were mapped to the orthologs of the pse-A and B genomes. Specifically, type I reads could only be mapped to the orthologs of the pse-A or B genomes. If one read was type I, we used the real number of the read to calculate RPM. Type II reads could be mapped to both the orthologs of the pse-A and B genomes; therefore, one particular read could be counted twice. If the read was considered a type II read, it was counted only once when calculating the RPM of a particular gene, because if it was counted twice, the methylation level would have been overestimated.

Identification of differentially methylated and expressed genes

The differentially expressed and methylated genes (t-test, $P < 0.05$) from peanut pod MEDIP-seq and RNA-seq datasets were identified based on RPM value using the software R (<http://bioconductor.org/packages/release/bioc/html/edgeR.html>). RPM was calculated as the clean read number mapped to the pse-A and B genomes of one gene per the total sequenced clean reads (millions). To identify the differentially expressed genes, we analyzed the RNA-seq data previously reported by Zhang (NCBI's Short Read Archive under accession number SRP064700) [29]. The plant samples for this MEDIP-seq study were also same to previous study [29].

Identification of peanut siRNA and their target loci in the pse-a and B genomes

Small RNAs were analyzed using our previously reported data [30]. The plant samples for this MEDIP-seq study were also same to previous study [30]. First, ribosomal RNAs, transfer RNAs, snoRNAs, and snRNAs were filtered to identify the siRNA candidates and miRNAs [30]. The TargetFinder software was used to identify the target loci in the pse-A and B genomes [36].

Identification of DNA methylation related genes

The protein coding sequences of Arabidopsis RNA-directed DNA methylation (RdDM) pathway genes were

used as queries to identify peanut RdDM pathway genes using BLASTP (E-value $< 10^{-20}$). Sequences with the smallest E-value and the highest identities were considered homologs of peanut genes.

Results

Strategy development for peanut DNA methylation and RNA-seq data analysis

Cultivated peanut is allotetraploid (AABB, $4n = 4x = 40$), originating from a single hybridization event between the diploid A and B genomes from wild type species [35, 37–39]. Whole genome sequencing of the two ancestral species (*A. duranensis* and *A. ipaensis*) has been completed (<http://peanutbase.org/>) [40]. The cultivated peanut genome was sequenced but not assembled. The corrected median identities between cultivated peanut genome sequence reads and the pseudomolecules of *Arachis. duranensis* and *Arachis. ipaensis* were 98.36 and 99.96%, respectively [40]. In this study, we used *A. duranensis* and *A. ipaensis* as A and B sub-genome (pse-genomes) of cultivated peanut for sequence data analysis.

Using the pse-A and pse-B genomes, the MEDIP-seq generated a total of approximately 42,857,143 clean reads from all samples (Short Read Archive in NCBI under accession number SUB4324289). MEDIP-seq data were mapped on pse-A and B genomes (Fig. 1). We found that the distribution of reads was different in gene coding regions compared with the promoter (upstream 2000 bp regions of genes) and transcript terminal regions (TTRs; downstream 2000 bp regions of genes) both in pse-A and pse-B genomes (Fig. 2). The CpG island was hyper-methylated compared with the CpG island shore (upstream and downstream 2000 bp regions of CpG island) both in pse-A and pse-B sub-genomes (Fig. 3).

When MEDIP-seq reads were mapped directly on the pse-A genome and pse-B genome, respectively, we encountered a problem. Reads were repeatedly mapped to the regions that were highly conserved between pse-A and B genomes, because of the existence of a large number of orthologous genes across these two sub-genomes. If the total MEDIP reads were mapped only to the A genome or only to B genome, the mapping rate was about approximately 79% because of the A or B genome specific genes. If we combined the reads mapped specifically to the pse-A or B genomes and to both genomes, the mapping rate increased to 95%. Therefore, our novel method improved the mapping strategy.

In this study, the gene methylation level was evaluated by RPM. Because we considered the orthologous genes as one gene, we had to count all reads that mapped to the orthologous genes located in both the pse-A or B genomes. The length of the orthologous genes could be

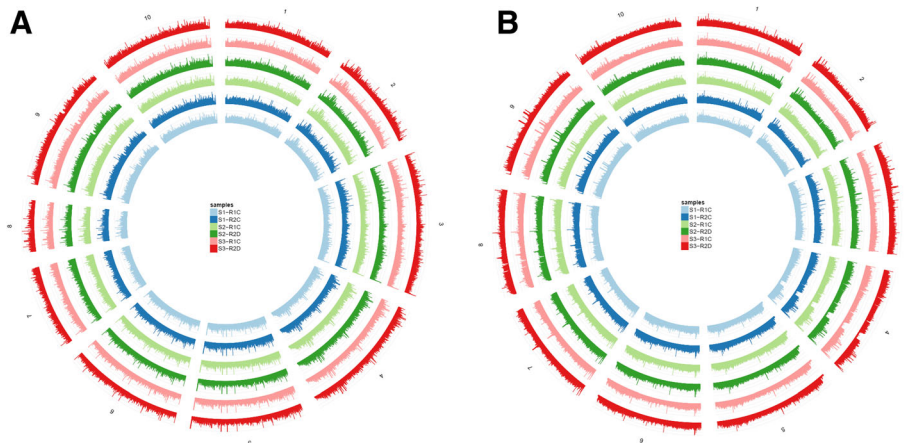


Fig. 1 Distribution of DNA methylation in S1, S2 and S3 A and B genomes. **a** represents A genome, and **b** represents B genome. The height of circular coordinate axis represents DNA methylation depth

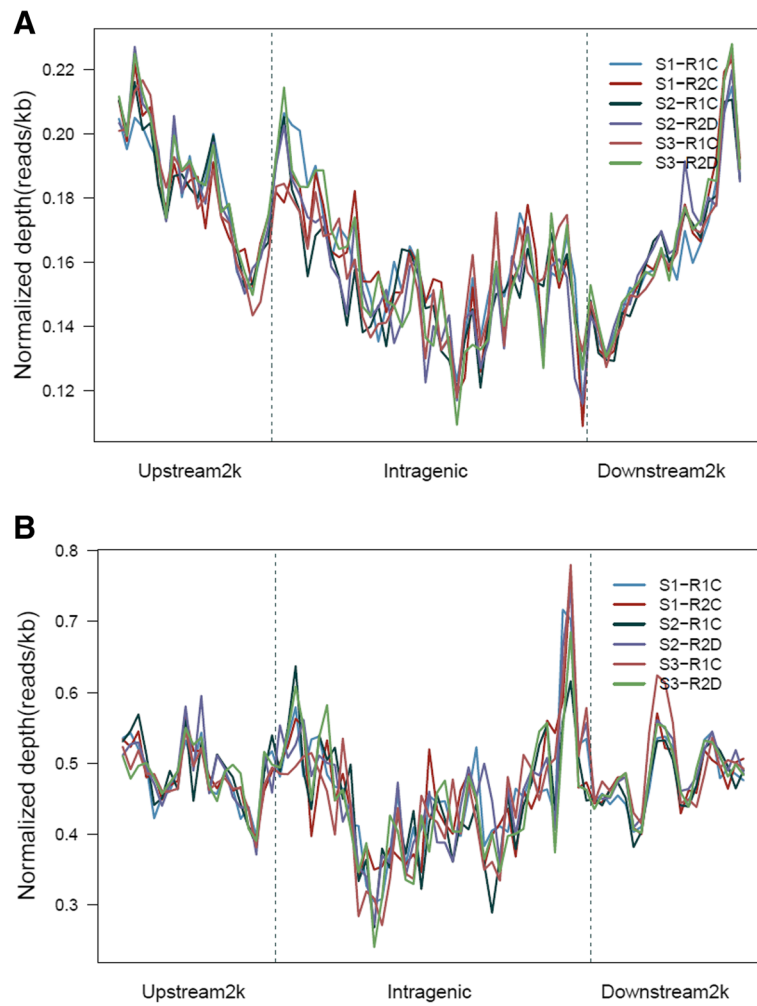


Fig. 2 Distribution of DNA methylation in S1, S2 and S3 genes from A and B genomes. **a** represents A genome, and **b** represents B genome. The height of coordinate axis represents DNA methylation depth

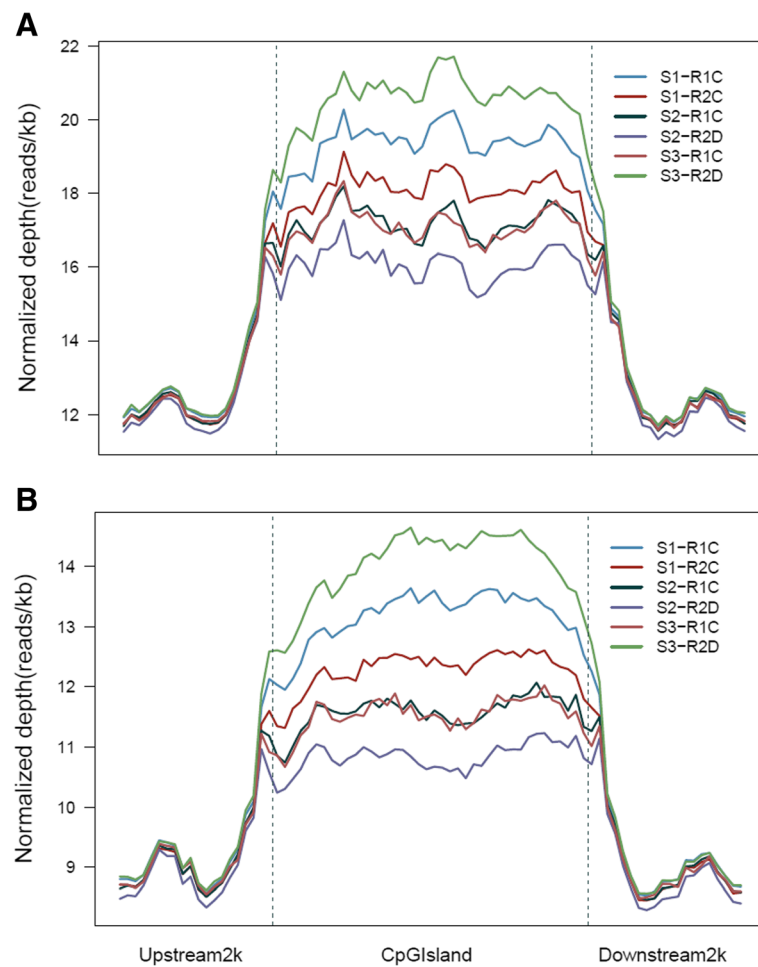


Fig. 3 Distribution of DNA methylation in S1, S2 and S3 CpG island and CpG island shore from A and B genomes. **a** represents A genome, and **b** represents B genome. The height of coordinate axis represents the DNA methylation depth

different; therefore, we could use RPKM, unlike studies conducted in rice [41].

This study focused on the DNA methylation of protein-coding genes. The orthologous genes across the pse-A and B genome were considered alleles, because we presumed that orthologous genes within the same genus are functionally similar. Functional annotation of orthologous genes across pse-A and B genomes are listed in Additional file 1: Table S1. These DNA methylation or expression levels of orthologous genes were difficult to distinguish between the pse-A and B genomes. Likewise, DNA methylation or expression levels were also difficult to distinguish between paternal alleles and maternal alleles. However, the sum of DNA methylation of orthologous genes from the pse-A and B genome were determined. The non-orthologous genes from the A and B genomes were analyzed using the common method. We identified 40,639 and 46,985 protein-coding genes in the pse-A and B genomes, respectively, and 19,030 orthologous genes across the

pse-A genome and B genomes (Additional file 1: Table S1; Fig. 4).

Differential DNA methylation in protein-coding genes

Differentially methylated genes (DMG) were identified across the S1, S2, and S3 stages of the gynophores. There were 1055 DMGs between S1 and S2 including a large number of photosynthesis-related genes such as photosystem I assembly protein, photosystem II CP43 chlorophyll apoprotein, photosystem II D1, light-harvesting chlorophyll B-binding protein, and chlorophyll synthase genes. Cell division protein, cell wall protein expansin 2, and cell number regulator-like protein genes were differentially methylated in all three stages. WUSCHEL-related homeobox (WOX), nodulin, and growth regulating factor (GRF) zinc finger protein genes were identified as DMGs. The methylation levels of two senescence-associated genes were decreased in S2 and two AUX/IAA protein genes were increased in S2 compared with S1. The methylation levels of three

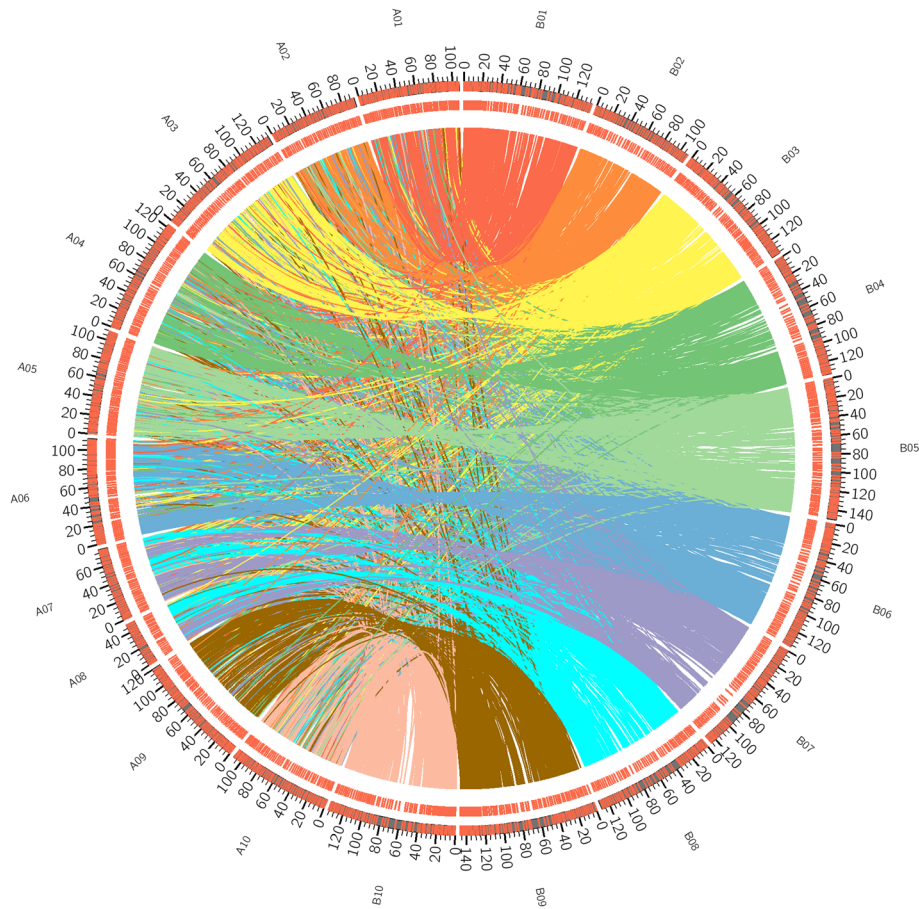


Fig. 4 Distribution of orthologous-pairs in A and B genomes. Orthologous-pair members from A and B genomes were linked by color lines. A01-A10 represents chromosome 1–10 in A genome and B01-B10 represents chromosome 1–10 in B genome

ethylene-responsive transcription factor genes were down-regulated in S2 (Additional file 2: Table S2).

In total, 1190 DMGs were identified between S1 and S3, including cell division FtsZ-like protein, auxin response factor, auxin-responsive family protein, and GRF zinc finger protein genes. The methylation levels of four senescence-associated genes were all up-regulated in S3 (Additional file 2: Table S2). DNA methylation of the DME gene was down-regulated in S3 compared with S1.

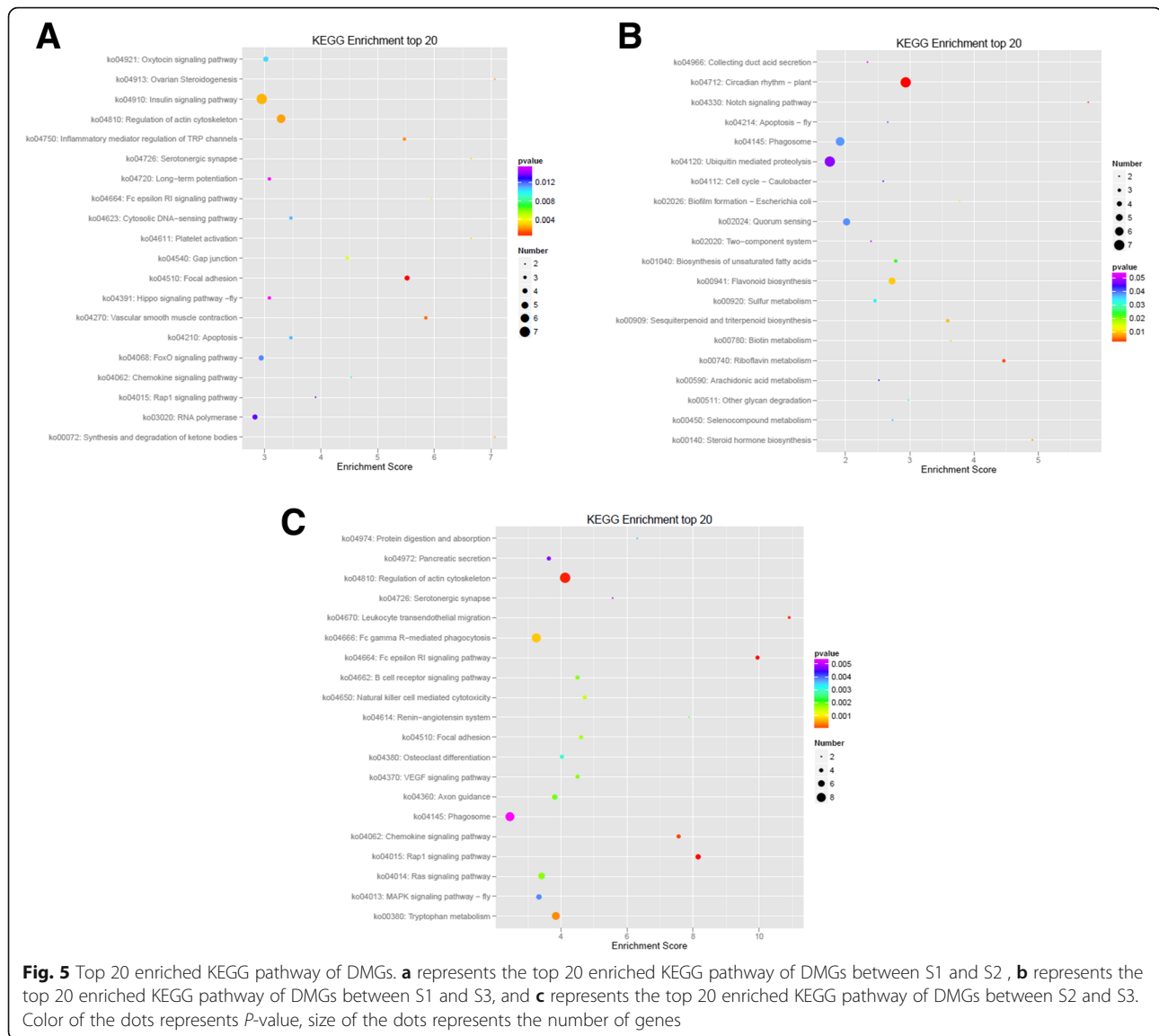
The number of DMGs between S2 and S3 was 2207. Embryo defective genes were found to be differentially methylated. Genes involved in cell division or expansion, such as cell division cycle protein, cell wall protein EXP2, and cell number regulator-like protein genes, were differentially methylated. Twelve GRF zinc finger protein genes and five nodulin genes were identified as DMGs between S2 and S3. The methylation level of the PIF1 gene was down-regulated while the methylation level of the PIF3 gene was up-regulated in S3. A number of genes involved in hormone signal transduction, such as ethylene responsive transcription factor and auxin response factor genes, were identified as DMGs. DNA

methylation levels of five senescence-associated genes increased in S3 (Additional file 2: Table S2).

KEGG pathway and enrichment analysis showed that DMGs were enriched ($P < 0.05$) in mRNA surveillance, oxidative phosphorylation, and regulation of actin cytoskeleton pathways between S1 and S2 (Fig. 5a; Additional file 3: Table S3). Some DMGs between S1 and S2 were involved in hormone signaling transduction. DMGs between S1 and S3 were enriched ($P < 0.05$) in ubiquitin mediated proteolysis, flavonoid biosynthesis, riboflavin metabolism, and sesquiterpenoid and triterpenoid biosynthesis (Fig. 5b; Additional file 3: Table S3). DMGs between S2 and S3 were enriched ($P < 0.05$) in regulation of actin cytoskeleton, RNA transport, cAMP signaling pathway, porphyrin and chlorophyll metabolism, tryptophan metabolism, and pentose phosphate pathway (Fig. 5c; Additional file 3: Table S3). Some DMGs between S2 and S3 were involved in hormone signaling transduction and cell cycle.

Mapping and analysis of RNA-seq data

Using the improved analysis strategy, we analyzed RNA-seq data previously published by Zhang (NCBI



accession number SRP064700) [29]. Results showed that the reads mapping rate of these sequences increased from 70 to 95% using our new strategy.

Many differentially expressed genes (DEGs) were found across the three developmental stages of gynophores. In total, 2337 DEGs were identified between gynophores in S1 and S2. Many photosynthesis-related genes were down-regulated in S2 including photosystem I reaction center subunit X, light-harvesting chlorophyll B-binding protein, and photosystem II reaction center protein genes. Two senescence-associated proteins, gibberellin 20 oxidase 1, and gibberellin 20 oxidase 2 genes were down-regulated in S2. Auxin response factor 4 and auxin transporter-like protein 5-like isoform X1 genes were also down-regulated in S2. Many ethylene-responsive transcription factor genes and three cell wall protein genes were up-regulated in S2 (Additional file 4: Table S4).

Between S1 and S3, 3169 DEGs were identified. Two differentially expressed ABA receptor genes were down-regulated in S3, and many hormone (ethylene, gibberellin, and auxin) signal transduction related genes were differentially expressed between S1 and S3. Many nodulin, nodulin-like, and WOX genes were differentially expressed between S1 and S3 (Additional file 4: Table S4). Seven nodulin genes were up-regulated in S3, and many WOX and hormone signal transduction related genes were differentially expressed between S1 and S3. Two differentially expressed ABA receptors were down-regulated in S3 (Additional file 4: Table S4). Between S2 and S3, 1849 DEGs were discovered.

Correlation between expression and DNA methylation

RdDM is the major small RNA-mediated epigenetic pathway in plants. The dynamic expression of genes in

the RdDM pathway was analyzed (Table 1). We found that the *AGO6* (Aradu.E98LA), *DRM2*, *MET1*, and *DEFECTIVE IN MERISTEM SILENCING 3* (*DMS3*) were all up-regulated in S2 compared with the expression in S1 (Fig. 6).

Compared with the expression in S1, *DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1* (*DRD1*), *MICRORCHIDIA 1* (*MORC1*), and *INVOLVED IN DE NOVO 2* (*IDN2*) were all up-regulated in S3 (Fig. 6).

LYSINE-SPECIFIC HISTONE DEMETHYLASE 1 (*LDL1*) was expressed at the highest level in S3 compared with that in S1 and S2. The expression of *LDL1* between S1 and S3 as well as S2 and S3 were different, while DNA methylation levels of this gene were different only between S1 and S2. DNA methylation levels of most genes in the RdDM pathway were stable across three stages except *NRPE5*, *HISTONE DEACETYLASE 6* (*HDA6*), *LDL1*, and *DMS3*. The expression levels of other demethylase genes were stable in the development of the gynophores from S1 to S3 (Fig. 6).

Differentially methylated and expressed genes (DMEGs)

Between S1 and S2, 69 DMEGs were detected, including photosystem I assembly protein, light-harvesting chlorophyll B-binding protein, and photosystem II protein D1 genes, which are all related to photosynthesis. One nodulin gene, one Dof zinc finger protein gene, and one ICE1-like transcription factor were DMEGs. Some DMEGs between S1 and S2 were involved in disease resistance such as disease-resistance response protein (Additional file 5: Table S5). DMEGs between S1 and S2 were enriched in stem cell fate determination, root development, lignin biosynthetic process, response to temperature stimulus, red light, blue light, and UV light (Additional file 6: Table S6).

Between S1 and S3, 117 DMEGs were detected, including photosystem I reaction center subunit IV, photosystem II CP43, and chlorophyll apoprotein genes. Some DMEGs between S1 and S3 were involved in hormone signal transduction, such as auxin-responsive family protein genes. AP2-like ethylene-responsive transcription factor and bHLH transcription factor genes were DMEGs between S1 and S3 (Additional file 5: Table S5). DMEGs between S1 and S3 were enriched in post-embryonic morphogenesis, organ boundary specification between lateral organs and meristem, auxin biosynthesis, and cell wall organization (Additional file 6: Table S6).

Between S2 and S3, 61 DMEGs were detected, including some photosynthesis related genes, germin-like protein, LRR, and NB-ARC domain disease resistance genes (Additional file 5: Table S5). DMEGs between S2 and S3 were enriched in embryo development, cold acclimation,

regulation of cell division, seed coat development, embryo sac development, and post-embryonic morphogenesis (Additional file 6: Table S6).

The relationship between gene DNA methylation and expression levels

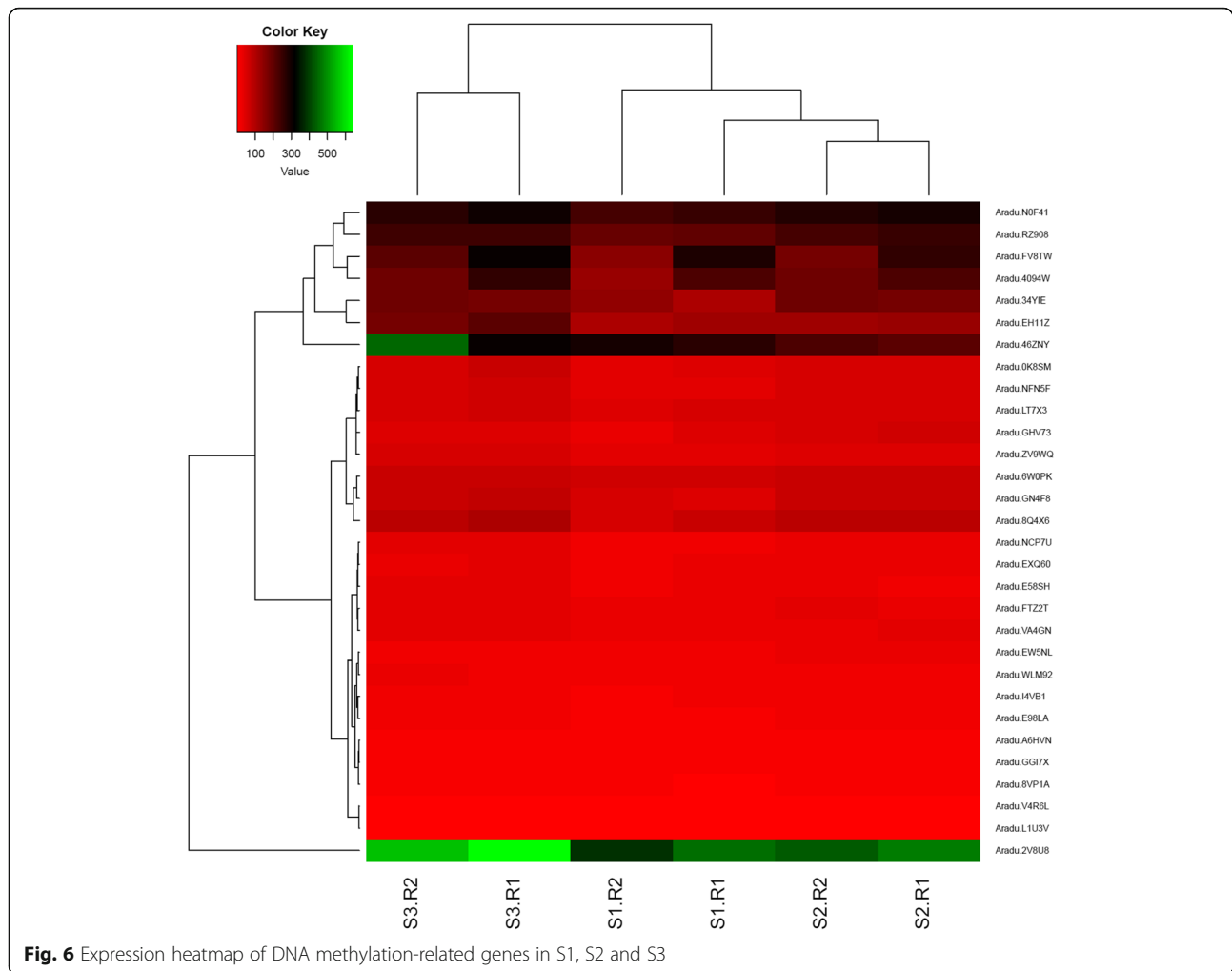
Most protein-coding genes were stable both in their expression and methylation across the three developmental stages. In the 69 identified DMEGs between S1 and S2, 24 were differentially methylated in the promoter region, 22 in the gene body, and 22 in the TTR. Among them, the expression of 29 genes (42.03%) was negatively related to their DNA methylation level. Ten of these 29 DMEGs were differentially methylated in the promoter region, eight in gene body, and 11 in the TTR (Additional file 5: Table S5). In 117 DMEGs between S1 and S3, 39 were differentially methylated in the promoter, 47 in the gene body, and 31 in the TTR. The expression of 47 genes (40.17%) was negatively related to their DNA methylation level. Among these 47 DMEGs, 11 genes were differentially methylated in the promoter, 26 in the gene body, and 10 in the TTR (Additional file 5: Table S5). In the 61 DMEGs between S2 and S3, 24 were differentially methylated in the promoter, 17 in the gene body, and 20 in TTR. The expression of 21 genes (34.43%) was negatively related to the DNA methylation level. Among these 21 DMEGs, nine genes were differentially methylated in the promoter, six in the gene body, and six in the TTR (Additional file 5: Table S5).

The relationship between small RNA level and DNA methylation

We analyzed the relationship between the abundance of 24-nucleotide (nt) siRNAs and the DNA methylation level of their target loci in peanut gynophores using previously reported data from small RNA in three stages of gynophore development [30] (Additional file 7: Table S7). The abundance of some 24-nt siRNAs was indeed positively associated with the DNA methylation level of some target loci such as siRNA t0207414 and t0470617. Twenty 24-nt siRNAs were not expressed in S3 but expressed in S1, and 30 of their target loci DNA methylation levels were decreased in S3. Thirteen 24-nt siRNAs were not expressed in S2 but were expressed in S1, and 45 of their target loci had decreased DNA methylation levels in S2. The relationship between the abundance of miRNA and DNA methylation level of their target loci was investigated using our previously reported miRNA data [30]. A positive relationship was observed between the abundance of some miRNAs and target loci methylation such as peanut miR5713, miR2108b, and miR8762d.

Table 1 DNA methylation-related genes in peanut

Gene name	Orthologous from A genome	Orthologous from B genome	Orthologs in Arabidopsis
NUCLEAR RNA POLYMERASE D1(NRPD1)	Aradu.EXQ60	Araip.NHV7F	AT1G63020
NRPE1	Aradu.U7301	Araip.A5H9W	AT2G40030
NRPD2/NRPE2	Aradu.8Q4X6	Araip.AVR8T	AT3G23780
NRPD4/NRPE4	Aradu.WLM92	Araip.F0FD4	AT4G15950
NRPE5	Aradu.LT7X3	Araip.2H1ED	AT3G57080
NRPE9B	Aradu.8VP1A	Araip.4786 J	AT4G16265
NRPB1	Aradu.FV8TW	Araip.AAM5W	AT4G35800
RNA-DEPENDENT RNA POLYMERASE 2 (RDR2)	Aradu.0K8SM	Araip.63JS7	AT4G11130
DICER-LIKE 3 (DCL3)	Aradu.EH11Z	Araip.0XV9H	AT3G43920
HUA ENHANCER 1 (HEN1)	Aradu.I4VB1	Araip.0P57Y	AT4G20910
ARGONAUTE 4 (AGO4)	Aradu.2V8U8	Araip.DM5GK	AT2G27040
AGO6	Aradu.E98LA	Araip.70L1G	AT2G32940
AGO9	Aradu.2V8U8	Araip.DM5GK	AT5G21150
CLASSY 1 (CLSY1)	Aradu.TP9JP	Araip.VW316	AT3G42670
DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1 (DRD1)	Aradu.E58SH	Araip.VZ0N	AT2G16390
DEFECTIVE IN MERISTEM SILENCING 3 (DMS3)	Aradu.NCP7U	Araip.X02G0	AT3G49250
RNA-DIRECTED DNA METHYLATION 1 (RDM1)	Aradu.0VT9B	Araip.6D36H	AT3G22680
KOW DOMAIN-CONTAINING TRANSCRIPTION FACTOR 1 (KTF1)	Aradu.4094 W	Araip.J0JMG	AT5G04290
INVOLVED IN DE NOVO 2 (IDN2)	Aradu.RZ908	Araip.15C6I	AT3G48670
IDN2 PARALOGUE 1 (IDP1)	Aradu.N0F41	Araip.47FQ5	AT1G15910
IDP2	Aradu.N0F41	Araip.47FQ5	AT4G00380
DMS4	Aradu.EW5NL	Araip.C9F8X	AT2G30280
DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2)	Aradu.NFN5F	Araip.B5M84	AT5G14620
SUVH2	Aradu.A6HVN	Araip.M3BFQ	AT2G33290
SUVH9	Aradu.A6HVN	Araip.M3BFQ	AT4G13460
SUVR2	Aradu.VA4GN	Araip.1D0HT	AT5G43990
MICRORCHIDIA 1 (MORC1)	Aradu.FTZ2T	Araip.B1U9G	AT4G36290
MORC6	Aradu.FTZ2T	Araip.B1U9G	AT1G19100
SAWADEE HOMEODOMAIN HOMOLOGUE 1 (SHH1)	Aradu.GG17X	Araip.U8UTV	AT1G15215
HISTONE DEACETYLASE 6 (HDA6)	Aradu.L1U3V	Araip.WB197	AT5G63110
JUMONJI 14 (JM14)	Aradu.V4R6L	Araip.Z2DN9	AT4G20400
LYSINE-SPECIFIC HISTONE DEMETHYLASE 1 (LDL1)	Aradu.ZV9WQ	Araip.F5V5K	AT1G62830
LDL2	Aradu.9QT6G	Araip.M4152	AT3G13682
UBIQUITIN-SPECIFIC PROTEASE 26 (UBP26)	Aradu.A0F5T	Araip.R2ZRV	AT3G49600
NEEDED FOR RDR2.INDEPENDENT DNA METHYLATION (NERD)	Aradu.EB2 7 U	Araip.FJ8IU	AT2G16485
CHROMOMETHYLASE 2 (CMT2)	Aradu.6W0PK	Araip.71XM6	AT4G19020
CMT3	Aradu.34YIE	Araip.527SE	AT1G69770
METHYLTRANSFERASE 1 (MET1)	Aradu.GN4F8	Araip.RYZ61	AT5G49160
SUVH4	Aradu.I17CD	Araip.A9GQY	AT5G13960
DECREASED DNA METHYLATION 1 (DDM1)	Aradu.LCP0L	Araip.5QZ4M	AT5G66750
CHROMOMETHYLASE 2 (CMT2)	Aradu.6W0PK	Araip.71XM6	AT4G19020
METHYLTRANSFERASE 1 (MET1)	Aradu.GN4F8	Araip.RYZ61	AT5G49160
DME	Aradu.GHV73	Araip.81XFD	AT5G04560



Discussion

In wheat, the methylome sequencing reads were mapped onto the diploid ancestor sub-genome A, B, and D individually [42–44], using SNPs to distinguish three sub-genomes [42]. When this strategy was used in peanut, we could only detect the methylated regions of DNA where there are SNPs and the general DNA methylation level of the sub-genome. This method missed DNA methylation information from genes with no SNPs between the sub-genomes. Additionally, SNPs only covered part of the genome, and many clean reads not located near the SNP regions were filtered out. Thus, to improve the coverage, we combined the reads mapped specifically to the pse-A or B genomes and to both genomes, increasing the mapping rate of each sample (both MEDIP-seq and RNA-seq data) to 95%. Here, a novel strategy was used. We considered the orthologs as one gene and counted all reads that mapped to the orthologs of the pse-A and B genomes for the calculation of RPM. Therefore, the DNA methylation levels of nearly all peanut protein-coding genes

could be examined. However, using this strategy, we could only use the RPM value to analyze the dynamic DNA methylation of the same gene in different samples. DNA methylation levels could not be compared between different genes.

In the development of gynophores from S1 to S3, the DNA methylation levels of many genes varied. From S1 to S2, the expression levels of two key methyltransferase genes, DRM2 and MET1, were up-regulated. The increased expression of these genes may lead to DNA methylation changes from S1 to S2 gynophores, which experienced environmental change from light to dark conditions. A previous study showed that light could affect DNA methylation of some genes, such as maize PEPC genes [45]. S1 and S2 gynophores grow in the light and dark, respectively, and the methylation levels of many genes changed during these two stages. However, the methylation level of peanut PEPC genes remained unchanged between S1 and S2 gynophores. Many genes involved in cell division, expansion, ethylene, and auxin signal transduction were considered

DMGs in S1 and S2 gynophores. During S2 and S3, many genes involved in cell division, expansion, senescence, nodulin, ethylene, and auxin signal transduction were also DMGs that may be associated with embryo and pod development. DNA methylation levels of many GRF transcription factor genes involved in fruit development [46] were higher in S3 compared to S2 gynophores. Interestingly, DNA methylation levels of PIF1 and PIF3 genes, key components of light signal transduction involved in hypocotyl elongation and other developmental processes [47], varied in S2 and S3. DNA methylation affected genes involved in multiple biological processes, suggesting that DNA methylation may play a key role in early peanut pod development.

Previous studies have shown that DNA methylation could repress gene transcription [45, 48]. DNA methylation could also be positively related to gene expression [49–51]. In our study, we found that the expression levels of many genes were positively or negatively related to DNA methylation levels, while the expression levels of many other genes were not correlated with their methylation levels. The expression of some auxin and ethylene-related genes were correlated with their DNA methylation levels, and both of these hormones play important roles in peanut pod development [29]. In addition, gibberellins play important roles in early peanut pod development [29]. Similarly, the expression levels of many gibberellin-related genes varied among S1, S2, and S3; however, the DNA methylation levels of these gibberellin-related genes were stable. A previous study showed that ABA was very important for early peanut pod development [29]. Our results showed that ABA-regulated genes were enriched in DMEGs of S1 and S3. In other words, the expression of many ABA-regulated genes was correlated with their methylation levels.

Walls are thin 1 (WAT1), an Arabidopsis homolog of nodulin MtN21, is required for secondary wall formation [52]. Nodulin MtN21 is involved in nodule development [53]. A previous study indicated that a senescence-associated nodulin gene is a candidate for embryo abortion of aerial pods [54]. The expression levels of many nodulin genes varied in S1, S2, and S3, and the expression levels of some nodulin genes were correlated with their methylation levels. A previous study showed that high expression of senescence-associated genes might lead to embryo abortion. For example, the gynophores growing in continuous light express a high level of senescence-associated genes and cannot form pods. Down-regulation of these genes under dark conditions could be a critical factor for the initiation of embryo and pod development [54]. Our RNA-seq results showed that one senescence-associated gene was down-regulated in S2 compared to S1. However, the expression of this gene was not correlated with its methylation level.

Small RNA is an important trigger of the RdDM pathway [24]. Small RNA could affect embryo or seed development by DNA methylation. For example, small RNA could be transported from the endosperm to the embryo leading to embryo hypermethylation in rice [55]. A previous study showed that 24-nt siRNAs play a major role in the RdDM pathway [24]. A recently conducted study showed that DNA methylation of targets depended on miRNAs [56]. Our results are in agreement with this recent study, confirming that the abundance of 24-nt siRNAs and miRNAs were indeed positively associated with DNA methylation of their target loci in peanut pods.

Conclusions

The improved mapping strategy could efficiently map reads to allotetraploid peanut DNA, increasing the mapping rate from 79 to 95%. MeDIP-seq data indicated that the methylation varied among the S1, S2, and S3 gynophores. The methylation levels of some genes differed in S1 to S3, which could be associated with the expression level of these genes. Genes involved in cell division, expansion, senescence, nodulin, ethylene, and auxin signal transduction were among the DMGs and may be associated with embryo development and pod formation. The expression levels of many genes involved in pod expansion varied in S1, S2, and S3. The senescence-related and nodulin genes may play important roles in peanut pod development. The change of the methylome during early peanut pod development is useful for understanding the molecular mechanisms that regulate peanut pod development.

Additional files

Additional file 1: Table S1. Orthologous genes from peanut A and B sub-genome. (XLS 3611 kb)

Additional file 2: Table S2. Differential DNA methylation in protein-coding genes. (XLS 8009 kb)

Additional file 3: Table S3. Enriched KEGG pathway of DMGs. (XLS 35 kb)

Additional file 4: Table S4. Differentially expressed genes. (XLS 8234 kb)

Additional file 5: Table S5. Differentially methylated and expressed genes. The type “Exp” represent information of gene expression, and “Meth” represent information of gene DNA methylation. (XLS 115 kb)

Additional file 6: Table S6. Enriched GO term of DMEGs. (XLS 86 kb)

Additional file 7: Table S7. siRNA expressed in three stages. (XLS 1247 kb)

Abbreviations

ABA: Abscisic acid; bHLH: Basic helix-loop-helix; CMT: Chromomethylase; DEG: Differentially expressed gene; DME: Demeter; DMEG: Differentially methylated and expressed gene; DMG: Differentially methylated genes; DMS3: DEFECTIVE IN MERISTEM SILENCING 3; DRD1: DEFECTIVE IN RNA-DIRECTED DNA METHYLATION 1; DRM: Domains Rearranged Methyltransferase; GA: Gibberellin; IDN2: INVOLVED IN DE NOVO 2; LDL1: LYSINE-SPECIFIC HISTONE DEMETHYLASE 1; MET: Methyltransferase; MORC1: MICRORCHIDIA 1; RdDM: RNA-directed DNA methylation; ROS1: Repressor of silencing 1; RPM: MEDIP-seq reads of gene per million mapped reads

Acknowledgements

Not applicable.

Funding

This work was financially supported by National Natural Science Foundation of China (31500217, 31500257, 31471526), Shandong provincial crop elite variety development project (2016LZGC025), Agricultural scientific and technological innovation project of Shandong Academy of Agricultural Sciences (CXGC2018E17) and Agricultural scientific and technological innovation project of Shandong Academy of Agricultural Sciences-cultivating project for National Natural Science Foundation of China in 2018.

Availability of data and materials

The data sets supporting the results of this article are included within the article and its additional files.

Authors' contributions

PW, HX and XW designed the study. PW wrote the manuscript and XW and HX revised the manuscript. SS, JM, YZ, HS, CG, CZ, SZ, LJ, SF and LH carried out the experiment, data analysis, and wrote the method section of the manuscript. All authors read and approved the final version of the manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Biotechnology 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.

²Shandong Academy of Grape, Jinan 250100, People's Republic of China.

³Life Science College of Shandong Normal University, Jinan 250014, People's Republic of China. ⁴Life Science College of Shandong University, Jinan 250100, People's Republic of China. ⁵Instituto Tecnológico del Valle de Oaxaca, 68000 Oaxaca, Mexico.

Received: 15 February 2018 Accepted: 20 November 2018

Published online: 13 December 2018

References

- Moctezuma E, Feldman LJ. Growth rates and auxin effects in graviresponding gynophores of the peanut, *Arachis hypogaea* (Fabaceae). *Am J Bot*. 1998;85:1369–76.
- Moctezuma E, Feldman LJ. The role of amyloplasts during gravity perception in gynophores of the peanut plant (*Arachis hypogaea*). *Ann Bot*. 1999;84:709–14.
- Xing S, Xi X. An ultrastructural observation of endosperm cellularization in *Arachis hypogaea*. *Acta Bot Yunnanica*. 1995;17:433–8.
- Xia H, Zhao C, Hou L, et al. Transcriptome profiling of peanut gynophores revealed global reprogramming of gene expression during early pod development in darkness. *BMC Genomics*. 2013;14:517.
- Moctezuma E. The peanut gynophore: a developmental and physiological perspective. *Can J Bot*. 2003;81:183–90.
- Periasamy K, Sampooram C. The morphology and anatomy of ovule and fruit development in *Arachis hypogaea* L. *Ann Bot*. 1984;53:399–411.
- Olsen O. Nuclear endosperm development in cereals and *Arabidopsis thaliana*. *Plant Cell*. 2004;16:S214–27.
- Thompson GE, Ziv M, Deitzer GF. Photocontrol of peanut (*Arachis hypogaea* L.) embryo and ovule development in vitro. *Plant Physiol*. 1985;78:370–3.
- Thompson LK, Burgess CL, Skinner EN. Localization of phytochrome during peanut (*Arachis hypogaea* L.) gynophore and ovule development. *Am J Bot*. 1992;79:828–32.
- Seo M, Nambara E, Choi G, Yamaguchi S. Interaction of light and hormone signals in germinating seeds. *Plant Mol Biol*. 2009;69:463–72.
- Gallego-Bartolome J, Minguet G, Grau-Enguix F, et al. Molecular mechanism for the interaction between gibberellin and brassinosteroid signaling pathways in *Arabidopsis*. *PNAS*. 2012;109:13446–51.
- Zhong S, Shi H, Xue C, et al. A molecular framework of light-controlled phytohormone action in *Arabidopsis*. *Curr Biol*. 2012;22:1530–5.
- Zhong X, Du J, Hale CJ, et al. Molecular mechanism of action of plant DRM de novo DNA methyltransferases. *Cell*. 2014;157:1050–60.
- Klose RJ, Bird AP. Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci*. 2006;31:89–97.
- Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. *Nature*. 2007;447:433–40.
- Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet*. 2010;11:204–20.
- Serman A, Vlahović M, Serman L, Bulić-Jakus F. DNA methylation as a regulatory mechanism for gene expression in mammals. *Coll Antropol*. 2006;30:665–71.
- Hsieh TF, Ibarra CA, Silva P, et al. Genome-wide demethylation of *Arabidopsis* endosperm. *Science*. 2009;324:1451–4.
- Rai K, Huggins IJ, James SR, Karpf AR, Jones DA, Cairns BR. DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell*. 2008;135:201–1212.
- Garg R, Kumari R, Tiwari S, Goyal S. Genomic survey, gene expression analysis and structural modeling suggest diverse roles of DNA methyltransferases in legumes. *PLoS One*. 2014;9:e88947.
- Du J, Zhong X, Bernatavichute YV, et al. Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell*. 2012;151:167–80.
- Henderson IR, Deleris A, Wong W, et al. The de novo cytosine methyltransferase DRM2 requires intact UBA domains and a catalytically mutated paralog DRM3 during RNA-directed DNA methylation in *Arabidopsis thaliana*. *PLoS Genet*. 2010;6:e1001182.
- Noy-Malka C, Yaari R, Itzhaki R, et al. A single CMT methyltransferase homolog is involved in CHG DNA methylation and development of *Physcomitrella patens*. *Plant Mol Biol*. 2014;84:719–35.
- Matzke MA, Moshier RA. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet*. 2014;15:394–408.
- Penterman J, Zilberman D, Huh JH, Ballinger T, Henikoff S, Fischer RL. DNA demethylation in the *Arabidopsis* genome. *PNAS*. 2007;104:6752–257.
- Zhu JK. Active DNA demethylation mediated by DNA glycosylases. *Annu Rev Genet*. 2009;43:143–66.
- Xiao W, Custard KD, Brown RC, et al. DNA methylation is critical for *Arabidopsis* embryogenesis and seed viability. *Plant Cell*. 2006;18:805–14.
- Ecker JR. Epigenetic trigger for tomato ripening. *Nat Biotechnol*. 2013; 31:119–20.
- Zhang Y, Wang P, Xia H, et al. Comparative transcriptome analysis of basal and zygote-located tip regions of peanut ovaries provides insight into the mechanism of light regulation in peanut embryo and pod development. *BMC Genomics*. 2016;17:606.
- Gao C, Wang P, Zhao S, et al. Small RNA profiling and degradome analysis reveal regulation of microRNA in peanut embryogenesis and early pod development. *BMC Genomics*. 2017;18:220.
- Wang P, Xia H, Zhang Y, et al. Genome-wide high-resolution mapping of DNA methylation identifies epigenetic variation across embryo and endosperm in maize (*Zea mays*). *BMC Genomics*. 2015;16:21.
- Langmead B, Cole T, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10:R25.
- Sato S, Nakamura Y, Kaneko T, et al. Genome structure of the legume, *Lotus japonicus*. *Nat Res*. 2008;15:227–39.
- Lin Y, Cheng Y, Jin J, et al. Genome duplication and gene loss affect the evolution of heat shock transcription factor genes in legumes. *PLoS One*. 2014;9:e102825.
- Wang P, Gao C, Bian X, et al. Genome-wide identification and comparative analysis of cytosine-5 DNA methyltransferase and demethylase families in wild and cultivated peanut. *Front Plant Sci*. 2016;7:7.
- Fahlgren N, Carrington JC. miRNA target prediction in plants. *Methods Mol Biol*. 2010. https://doi.org/10.1007/978-1-60327-005-2_4.

37. Kochert G, Stalker HT, Gimenes M, Galgaro L, Romero-Lopes C, Moore K. RFLP and cytogenetic evidence on the origin and evolution of allotetraploid domesticated peanut, *Arachis hypogaea* (Leguminosae). *Am J Bot.* 1996;83:1282–91.
38. Freitas FO, Moretzsohn MC, Valls JF. Genetic variability of Brazilian Indian landraces of *Arachis hypogaea* L. *Genet Mol Res.* 2007;6:675–84.
39. Moretzsohn MC, Gouvea EG, Inglis PW, Leal-Bertioli SCM, Valls JFM, Bertioli DJ. 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.* 2013;111:113–26.
40. Bertioli DJ, Cannon SB, Froenicke L, et al. The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat Genet.* 2016;48:438–46.
41. Xing MQ, Zhang YJ, Zhou SR, et al. Global analysis reveals the crucial Roles of DNA methylation during Rice seed development. *Plant Physiol.* 2015;168:1417–32.
42. Gardiner LJ, Quinton-Tulloch M, Olohan L, Price J, Hall N, Hall A. A genome-wide survey of DNA methylation in hexaploid wheat. *Genome Biol.* 2015;16:273.
43. Brechley R, Spannag M, Pfeifer M, et al. Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature.* 2012;491:705–10.
44. Wilkinson PA, Winfield MO, Barker GL, et al. CerealsDB 2.0: an integrated resource for plant breeders and scientists. *BMC Bioinformatics.* 2012;13:219.
45. Tolley BJ, Woodfield H. Light-regulated and cell-specific methylation of the maize PEPC promoter. *J Exp Bot.* 2012;63:1381–90.
46. Liu X, Guo LX, Jin LF, et al. Identification and transcript profiles of citrus growth-regulating factor genes involved in the regulation of leaf and fruit development. *Mol Biol Rep.* 2016;43:1059–67.
47. Li K, Yu R, Fan LM, Wei N, Chen H, Deng XW. DELLA-mediated PIF degradation contributes to coordination of light and gibberellin signalling in *Arabidopsis*. *Nat Commun.* 2016;7:11868.
48. Li X, Zhu JD, Hu FY, et al. Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. *BMC Genomics.* 2012;13:300.
49. Song Q, Guan X, Chen ZJ. Dynamic roles for small RNAs and DNA methylation during ovule and Fiber development in Allotetraploid cotton. *PLoS Genet.* 2015;11:e1005724.
50. Tran RK, Zilberman D, Bustos CD, et al. Chromatin and siRNA pathways cooperate to maintain DNA methylation of small transposable elements in *Arabidopsis*. *Genome Biol.* 2005;6:R90.
51. Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat Genet.* 2007;39:61–9.
52. Ranocha P, Denancé N, Vanholme R, et al. Walls are thin 1 (WAT1), an *Arabidopsis* homolog of *Medicago truncatula* NODULIN21, is a tonoplast-localized protein required for secondary wall formation in fibers. *Plant J.* 2010;63:469–83.
53. Heller G, Lundén K, Finlay RD, Asiegbo FO, Elfstrand M. Expression analysis of Clavata1-like and Nodulin21-like genes from *Pinus sylvestris* during ectomycorrhiza formation. *Mycorrhiza.* 2012;22:271–7.
54. Zhu W, Chen X, Li H, et al. Comparative transcriptome analysis of aerial and subterranean pods development provides insights into seed abortion in peanut. *Plant Mol Biol.* 2014;85:395–409.
55. Zemach A, Kim MY, Silva P, et al. Local DNA hypomethylation activates genes in rice endosperm. *PNAS.* 2010;107:18729–34.
56. Wu L, Zhou H, Zhang Q, et al. DNA methylation mediated by a microRNA pathway. *Mol Cell.* 2010;38:465–75.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

