Single molecular real-time sequencing revealing the ovule abortion regulatory mechanisms in the female-sterile line of *Pinus tabuliformis* Carr.

Zaixin Gong

Beijing Forestry University

Rui Han

Beijing Forestry University

Min Zhang

Beijing Forestry University

Qianquan Yang

Beijing Forestry University

Ming Zeng

Beijing Forestry University

Yuanyuan Zhao (✉ yyzhao@bjfu.edu.cn)

Beijing Forestry University, Beijing Advanced Innovation Center for Tree Breeding by Molecular Design

Caixia Zheng

Beijing Forestry University

Research article

Keywords: *Pinus tabuliformis* Carr., ovule abortion, single molecule real-time sequencing, next-generation sequencing, alternative splicing, simple sequence repeat

Posted Date: September 29th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-51315/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

Ovule abortion is a common phenomenon in plants, which has an essential significance in seed production. The development of the female gametophyte (FG) is one of the crucial processes of the life cycle, and FG dysplasia is a significant cause contribute to ovule abortion in many species. However, because it is difficult to acquire the mutant about ovule development in gymnosperms, previous studies of the ovule and FG development are mainly focused on angiosperms, especially in *Arabidopsis thaliana*. Thus the investigation on this field of gymnosperms remains unclear.

Results

In this study, we investigated the transcriptome data of wild-type (female fertile line, FER) and natural ovule abortion mutant (female sterile line, STE) of *Pinus tabulaeformis* Carr. to evaluate the mechanism of ovule abortion during the process of free nuclear mitosis (FNM). Using single molecule real-time (SMRT) sequencing and next-generation sequencing (NGS), we obtained 202,869 (FER), 197,977 (STE) mapped full-length non-chimeric (FLNC) reads and analyzed 18 cDNA libraries. Based on the SMRT sequencing, we found that both GO annotation and KEGG pathway terms results were similar in FER and STE. NGS analysis further showed altogether 99 differentially expressed genes (DEGs) with opposite expression patterns during the FNM process in FER and STE. According to SMRT sequencing, we found that the number of isoforms and alternative splicing (AS) patterns are variable between FER and STE. Moreover, 5,530 and 5,096 potential simple sequence repeats (SSRs) were identified in FER and STE, respectively. Functional annotation results demonstrated that genes involved in energy metabolism, signal transduction, cell division, and stress response were differentially expressed in different lines, 9 pairs of which were confirmed by quantitative real-time PCR.

Conclusion

Taken together, these results provide some new insights about the ovule abortion in gymnosperms and further reveal the regulatory mechanisms of ovule development.

Background

Sexual reproduction is the most evolved reproductive mode in plants, as well as a link in the life cycle of seed plants. The completion of sexual reproduction depends on the normal development of reproductive organs and embryos. The female sterility is a common phenomenon during sexual reproduction in the plant kingdom [1, 2], which might result in yield reduction, affect the fine variety breeding, as well as decrease the economic benefit of the cash crops and woody plants. Generally, female sterility may be induced by many factors, such as abnormal ovule development, the wrong number of polar nuclei, meiosis, and mitotic disorder. For example, female sterile mutants of *Punica granatum* L. is mainly caused by abnormal ovule development [3]. Awasthi et al. reported that unusual mitotic divisions with a reduced number of polar nuclei in megagametophytes led to female sterility in *Oryza sativa* L. [4]. Rosellini et al. found that incompleted meiosis resulted in embryo sacs lack and the accumulation of callose in the nucellus in female sterility lines of *Medicago sativa* L. [5]. In addition, the abnormal of female gametophyte formation was the main reason for ovule abortion, which was observed in female sterility lines of *Allium sativum* L. [6]. These studies suggested that the mechanism of female sterility is various and complex, thus, a better understanding of the molecular regulation of female sterility is essential for improving the germplasm resources and breeding the plus tree.

Previous studies revealed that many genes might be associated with sterility in plants. For example, *CLV1* regulates the number of carpels and seeds in *Arabidopsis* [7]. *SUS1* and *AGPL1* could affect the sucrose and starch metabolisms, which over-expression might accelerate the speed of vegetative and reproductive growth. Moreover, low expression perhaps eventually led to pollen sterility and increased the number of sterile seed in rice [8, 9]. In *Arabidopsis*, *AGO9* is highly expressed in the ovule and interacts with small RNAs derived from transposable elements (TEs); however, silence TEs in the female gametophyte

influenced the formation of megasporocyte [10, 11]. Cyclin genes regulate cell cycle progression and cell division. Huang indicated that the expression of cyclin genes affects the ovule development in *Arabidopsis* [12]. Kempe et al. found that the intron *UBQ10-i1* participates in the synthesis of functional cytotoxin by trans-splicing, and finally affects the male-sterile through pollen ablation in wheat [13]. Studies about plant sterile were mainly concentrated on herbs with short life cycle, while few researches of the woody plants on this field, particularly in gymnosperms.

Next-generation sequencing (NGS) is able to quantitative analyze transcripts correctly but the short read length limited the accuracy of gene structure prediction. Yet single molecule real-time (SMRT) sequencing such as the Pacific Biosciences platform (PacBio) could produce longer reads without assembly. Depend on the long reads of SMRT, datasets exhibit excellent at connecting different exons, up to entire transcripts [14, 15]. This methodology provides preferable condition for the analyze of alternative splicing (AS) and simple sequence repeats (SSRs) [16, 17]. Increasing evidence suggests that AS can change the structure of transcripts and proteins, when AS was activated in the initial regulatory period, it can control the entire developmental pathway effectively and regulate gene expression quantitatively [18, 19]. Several studies proved that AS is related to the vegetative growth of plants in many species, including *Arabidopsis thaliana* [20], *Salvia miltiorrhiza* [21], and *Astragalus membranaceus* [22]. Furthermore, Sun et al. indicated that the fifth intron retained caused the deletion of the sixth exon of *ohms1*, which even affected the male sterile in rice [23]. SSRs are tandem repeat nucleotides in DNA sequences, which can be found both in eukaryote and prokaryote. SSRs were often used as genetic markers, which also played important roles in the regulation of gene expression and the development of organism [24]. Dou et al. determined the location of the gene caused female sterile in wheat by SSRs marker [25]. Kato used SSRs marker to identify the position of *ST8*, which influence the male and female sterile in soybean [26]. However, there is no relevant research in gymnosperms.

Gymnosperms retain some characteristics of spore plants, and have the evolutionary characteristics of angiosperms. Therefore, as the node of plant evolution, gymnosperms has momentous research value, especially on the reproductive development [27]. *Pinus tabulaeformis* Carr., the main afforestation species in north China, is an endemic and evergreen tree species [28]. It is noteworthy that *P. tabulaeformis* is an excellent material for study of the molecular regulation of gymnosperm ovule development, because its ovule takes about 13 months to complete cellularization [29]. A previous study showed that the line No. 28, in a *P. tabulaeformis* seed orchard in Xingcheng, Liaoning, China, was a female-sterile mutated line. The appearance of the cone was normal but the free nuclear mitosis of megagametophyte (FNMM) was terminated at an early stage, which caused the ovule abortion, led to the abnormal production of seeds [30]. To understand the reasons of ovule abortion, our research team has analyzed above the platform of the NGS technologies and screened out a series of interested genes about mitosis and phytohormones [31]. Nevertheless, to further reveal the mechanisms of ovule development needs sufficient length of transcript reads, which is the most significant disadvantage of NGS.

In the present study, to extend our knowledge of the mechanism of *P. tabulaeformis* female sterility during the FNM process, we combined NGS and SMRT sequencing to generate a more complete full-length *P. tabulaeformis* transcriptome, which benefits for qualitative and quantitative analysis of the isoforms and AS events more accurately. Different development stages of fertile line ovules and sterile line ovules of *P. tabulaeformis*, were measured and compared. Combined sequencing was employed to identify transcripts and to separately examine their expression patterns in different lines, screen out potential differentially expressed genes (DEGs) participant in the regulation of ovule abortion, and the expression levels of these DEGs were validated by quantitative real-time PCR. In addition to this, the AS and SSRs of the mutant and normal ovules in *P. tabulaeformis* were explored according to the SMRT sequencing. Based on the these results, the molecular regulation of the FNMM half-stopping and ovule abortion in female sterile mutate line in *P. tabulaeformis* were further revealed, and our comprehensive analyses afford new insight into the molecular mechanisms of ovule development in gymnosperms.

Results

Combined sequencing approach to analyze the ovule of *P. tabulaeformis*

To investigate the process of FNM in ovule development in female fertile and sterile lines of *P. tabulaeformis*, reveal the reasons for the FNMM discontinue, we selected three different time points to further analyze the ovule development. Including the early

February (early period of FNM process), the early March (middle period of FNM process), and the early April (late period of FNM process). Female fertile line (FER) ovule samples at above periods were marked by FER-FNM1, FER-FNM2, and FER-FNM3, respectively, while female sterile line (STE) ovule samples were marked by STE-FNM1, STE-FNM2, and STE-FNM3, respectively, each sample in triplicate. Eighteen mRNA samples from different stages ovule samples were subjected to 2*150 paired-end sequencing using the HiSeq™ 4000 platform, with 961,064,004 reads produced (Fig. 1). The repeatability of the RNA-seq libraries was evaluated using a PCA analysis; the reproducibility of the libraries was good (Fig. S1). Then combined in equal amounts of total RNA from three stages (FNM1, FNM2, and FNM3). Finally, we got two pooled samples (FER and STE). Two samples were normalized and subjected to an SMRT sequencing using the PacBio RS II platform (Table S1). In total, 276,887 (FER) and 307,741 (STE) ROIs were generated respectively (Fig. 1), and most of them (73.27% in FER, 64.33% in STE) were FLNC reads, which the entire transcript region from the 5′ primer to the 3′ primer, and the 3′ poly(A) tail were observed (Fig. 2A), and the length of FLNC is longer in STE than FER (Fig. 2B, 2C).

As shown in Fig. 2F, there was a significant variation of transcript length distribution between two platforms, and the read length was longer by PacBio, only 1.5% of the transcripts from the PacBio reads were <600 bases, but nearly 60% of the assembled transcripts from NGS reads were in this range. Based on Illumina short-read data, the length of N50 was 1,767bp, whereas N50 was 3,564bp based on PacBio, demonstrated that data assembling by PacBio was better than Illumina. Therefore, NGS reads corrected using SMRT subreads is more valuable than relying on single platform data. Besides, combined sequencing is more worthwhile for the species without a reference genome.

Functional annotation of full-length transcriptomes of FER and STE ovule in *P. tabuliformis*

For annotation of genes, alignment searches were conducted against public bioinformatics databases. And to further evaluate the function distribution of our transcriptome, we used GO annotation to classify these genes in FER and STE, respectively (Fig. S2). Annotated genes were divided into three categories, biological process, cellular component, and molecular function. In the biological process category, both in FER and STE, annotated genes were significantly enriched in metabolic processes, cellular processes, and single-organism processes. Furthermore, cell, cell part, and organelle enriched more obviously in the cellular component in each line; in the molecular function category, the genes from different lines were enriched into similar processes, mainly for catalytic activity and binding.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) was also used to investigate the pathway and network of the molecular regulation of FER and STE. Both in FER and STE, annotated genes were mapped into 129 pathways (Table S2, S3). Same to the GO annotation, the classes of KEGG pathway between two lines terms were quite similar. Same in two lines, metabolic pathways, biosynthesis of secondary metabolites, biosynthesis of antibiotics and microbial metabolism in diverse environments were the top 4 pathways enriched the most abundant transcripts, but compared with FER, there were more genes enriched in these 4 pathways in STE (Table S2, S3). Although the number of genes in FER and STE enriched in each pathway was slightly different, however, there were no significant differences in pathway annotation between FER and STE.

We got a through functional characterization for the full length transcripts of the FER and STE ovule, relied on the results from GO and KEGG functional annotation and classification. The GO and KEGG results also indicated that the functional annotation and classification of transcripts between two lines were similar on the whole, showing the pathway level was conservative for the transcriptome of FER and STE in *P. tabuliformis*. However, the expression level of individual genes might be different, which might related to the ovule abortion.

Candidate genes with opposite expression patterns during the FNM process in STE and FER

According to the variety of gene expression during the whole development process in FER and STE ovules, genes could be clustered into eight profiles by Short Time-series Expression Miner (STEM) software. These model profiles have been chosen to summarize the expression pattern of the genes. Among the eight patterns, we identified four patterns of genes that shown very significant p-values (colored boxes) (Fig. S3), these four profiles are significant both in FER and STE, containing two down-regulated patterns (profile 1 and 0) and two up-regulated patterns (profile 6 and 7). Genes in profile 1 and 0 in STE have been combined, however, profile 6 and 7 have been added in FER. Then 23 DEGs up-regulated in FER also down-regulated in STE

(part 1) have been screened (Fig. 2D). 69 DEGs up-regulated in STE as well as down-regulated in FER (part 2) could be filtered equally (Fig. 2E). These genes expressed in these two sections established contrary trends in FER and STE during the FNM process, which could probably play a pivotal role in ovule development.

Two parts of genes were subjected to GO-term analysis, respectively. They were still classified into three main categories, including “cellular component”, “molecular function” and “biological process” (Fig. 2G, 2H). In the biological process category, metabolic process, cellular process, and single-organism process were the most significant in both two parts. In the cellular component category, cell and cell parts are more obvious. Most DEGs were in catalytic activity and binding two parts in the molecular function category. Then DEGs were subjected to KEGG pathway enrichment analysis. The KEGG pathways have been mapped in are shown in Table S4, S5. In part 1, annotated genes enriched in 14 KEGG pathways and 24 pathways in part 2.

Combining two parts of DEGs, Fig. 3 and Table S6 had listed the DEGs which expression level existed obvious differences in varying lines, that were likely associated with female sterility in *P. tabuliformis*. Energy metabolism and signal transduction are important for FNMM development. Therefore, we investigated the genes associated with carbohydrate metabolism and signal transduction, including sucrose synthase, phosphofructokinase, chorismate synthase, clavata 1-like protein, tetraspanin, reticulons and plasma membrane ATPase related genes. Among these DEGs, most showed significantly lower transcript levels in STE ovules. Specifically, the transcript levels of sucrose synthase genes *SUS* (Iso_0064722, Iso_0065236, Iso_0045687), *AGPL1* (Iso_0015412), clavata 1-like protein related genes *CLV1* (Iso_0025308, Iso_0025527, Iso_0073847, Iso_0079334), and some other transmembrane protein related genes *RTNLB8* (Iso_0006259), *TET18* (Iso_0006259), *NPF3.1* (Iso_0006259) were markedly lower in STE compared to FER.

There were also several DEGs related to mitosis and apoptosis, encoding cyclin (*CYCB*, Iso_0016253), tubulin (*TUBA*, Iso_0012780), dihydrofolate reductase (*DHFR*, Iso_0053543) and cysteine proteinase inhibitor (*CPI*, Iso_0031083), which differentially expressed between FER and STE also might lead to FNM half-stop. DEGs associated with mitosis like *CYCB*, *TUBA* and *DHFR* all kept a high expression level in FER and low in STE. However, *CPI* was high expressed in FER, but the low expression level of *CPI* maybe result in programmed cell death and active oxygen accumulated [32]. Similarly, the expression of *bHLH66* (Iso_0067886) was obviously different between FER and STE. Furthermore, the transcript levels of several genes related to the stress response, such as *PER* (Iso_0053653, Iso_0049597, Iso_0010250), *SOBIR1* (Iso_0024626), *exgA* (Iso_0058895) and *CALS3* (Iso_0078524) showed opposite expression levels in varying lines.

AS analysis of DEGs with different expression patterns during the FNM process in two lines

Describe the complexity of AS at the whole transcriptome scale is one advantage of full-length sequencing. To analyze the AS events of transcript isoforms, we partition transcripts into gene families, then reconstructed into a coding reference genome based on k-mer similarity (Fig. 4A). A total of 42.69% had one isoform in FER and a slightly lower percentage (36.91%) in STE. However, there were more contigs with more than ten isoforms in FER (4.49%, 653) compared with STE (4.44%, 608) (Fig. 4C). Seven AS modes (skipping exon, mutually exclusive exons, alternative 5' splice-site, alternative 3' splice-site, retained intron, alternative first exon, and alternative last exon) existed in the ovule of *P. tabuliformis* (Fig. 4B). In total, 1,830 and 1,243 AS events were identified in FER and STE, respectively. The retained intron was the most significant AS mode, which also the predominant mode in plants [33]. Together with alternative 5' and 3' splice-site AS events, these three types of AS event accounted over 80% of detected events. On the contrary, mutually exclusive exons was the least AS event (Fig. 4D).

To explore whether AS exists in the genes might affect female sterility in *P. tabuliformis*, we researched AS in the DEGs, which analyzed above with opposite expression patterns during the FNM process in varying lines. Finally, isoforms of *AGPL1* (associated with starch metabolism), *bHLH66* (associated with plant growth and development), and *TUBA* (associated with mitosis) were diverse in FER and STE (Table S7), RI was the main AS event and existed in all three DEGs, *TUBA* also had A3 splicing (Fig. 4E, S4). The number of isoforms was varied between FER and STE in *AGPL1* and *TUBA*, but *bHLH* had only one isoform in each line, however, the splicing isoforms were different.

Development and characterization of SSR marker

To further investigate the differences of regulatory mechanisms between FER and STE ovules, thus the microsatellites were identified in this study. 40,828 (in FER) and 44,195 (in STE) genes were generated to discover potential microsatellites and defined as di- to hexa-nucleotide motifs in this study. Using the MicroSatellite (MISA, <http://pgrc.ipk-gatersleben.de/misa/>), a total of 5,530 potential simple sequence repeats (SSRs) were identified in 4,276 genes in FER. Similarly, 5,096 SSRs were identified from 3,948 genes in the STE ovule. Of the 4,276 genes, 868 contained more than one SSR in FER, however, in STE, this ratio is 796 of 3,948 (Table 1).

In the present study, the characterizations of the potential 5,530 SSRs in FER and 5,096 in STE were further analyzed. As shown in Table 1, the di-nucleotide repeats (45.5% in FER, 49% in STE) ranked first, and next came the tri- (38.4% in FER, 36.4% in STE), tetra- (8.4% in FER, 8% in STE), hexa- (5.2% in FER, 4.5% in STE) and penta- nucleotide (2.4% in FER, 2.6% in STE) repeats. Besides, the number of repeat units from Di- to hexa-nucleotide motifs were further summarized. As shown in Table S8 and Fig. 5A, 5B, the mostly represented repeat units of potential SSRs was 4-7, which accounts for 74.8% in FER, and 72.2% in STE. Tri-nucleotide repeats was the most abundant in this part, both in FER and STE. FER has more potential SSRs than STE, and the quantity difference of 4-7 repeat unit between FER and STE was more significant than the other levels of repeat unit. From Fig. 5C, 5D, we found that the SSR motifs in two lines shared significant similarity. The AT/TA di-nucleotide repeat was the most abundant motif in FER (30.4%) and STE (32.1%), followed by AG/CT (11% FER, 12.1% STE), AGC/CTG (8.9% FER, 6.7% STE), AAG/CTT (7.5% FER, 7% STE), AAT/ATT (6% FER, 6.7% STE) and ATC/ATG (5.9% FER, 5.6% STE). These six types of nucleotide repeats mentioned above represented about 70%.

Verification of the Gene Expression Profile by qRT-PCR

To confirm the transcriptomic analysis results, DEGs were randomly selected for quantitative real-time PCR (qRT-PCR) validation using the same type of samples compared with formerly used samples in NGS analysis, including *SUS*, *PER12*, *PFK2*, *CLV1*, *SOBIR*, *UBQ10*, *TUBA*, *AGPL1*, and *TET18*. The primers of these DEGs were listed in Table S9. The expression profiles of the candidate DEGs revealed by qRT-PCR data were consistent with those derived from sequencing. As shown in Fig. 6, the energy metabolism related genes *SUS*, *AGPL1* and *PFK2* were high expressed in FER, the expression level of *PFK2* was significantly high in FNM3 stage, *SUS* and *AGPL1* were high in FNM1. The expression of *CLV1* and *TET18*, which is associated with signal transduction, and *TUBA* were markedly lower in the STE ovules compared with FER ovules. In addition, the expression levels of *PER12* and *UBQ10* that related to stress response were higher in FER than STE, especially in FNM1 stage. On the contrary, *SOBIR1* was obviously up-regulated in STE, which is associated with apoptosis. The expression profiles of the candidate DEGs revealed by qRT-PCR data were consistent with those derived from sequencing.

Discussion

The development of the FG is one of the key processes of alternation of generations for gametophyte and sporophyte in plant, which decides the fate of the seeds in the future. To date, plenty of studies of FG development during FNM is based on angiosperms, especially in *Arabidopsis* [34]. Few reports focused on the female sterile in woody plants, such as *Ulmus minor* Mill. [35] and *Xanthoceras sorbifolium* Bunge [2]. However, due to the ovule mutants of gymnosperms are rare and the ovules are small and difficult to obtain, the molecular regulation of ovule development remains largely unknown in gymnosperms. In this investigation, we combined NGS and SMRT to analyze the transcripts in different periods of varying lines in *P. tabuliformis* (Fig. 1). Previous studies about female sterile are mainly using NGS [36, 37], however, NGS was limited for the short length of reads, the sequencing products were incomplete and required assembly. SMRT sequencing can overcome the disadvantage of NGS because of its long reads [18]. NGS reads were corrected by SMRT subreads, which helps qualitative and quantitatively analyze the transcriptome differences between FER and STE more accurately. A total of 142.8G nucleotides was generated based on the illumina platform, and 276,887 (FER) and 307,741 (STE) ROIs based on the PacBio RS platform (Fig. 1). We found that no matter the average length or the N50 length, it is obviously that the full length transcripts were much longer than that of the *de novo* assembled transcripts (Fig. 2F).

Energy metabolism regulates the vital movement at the whole plant level, which is a critical pathway during reproductive growth. Several regulatory genes of starch and sucrose metabolism and glycolysis have been identified associated with

reproductive development. *SUS* regulates the number of bud primordia and accelerating bud growth [38]. Kong et al. found that the expression level of *SUS* and *AGPL* influenced the number of seeds, which is significantly low expressed in male-fertile lines of rice [9]. *PFK2* encodes 6-phosphofructokinase, which is the most important element for the regulation of glycolysis [39, 40]. In the present study, many genes encoding components associated with starch and sucrose metabolism and glycolysis showed different expression patterns during ovule development in FER compared with STE. In the FER ovules *SUS* and *AGPL1* were significantly high expressed in the FNM1 stage, and the expression level of *PFK2* was increased from FNM1 to FNM3, but in STE they all low expressed. This result indicated that the accumulation and metabolism of starch and sucrose might be limited, and had obstacles in glucose catabolism. Therefore, substance accumulation of STE ovule maybe insufficient, unable to provide a suitable environment for the development of FG (Fig. 7).

Recent years, many genes related to inter- and intracellular transport which led to ovule abortion have been reported. Huck et al. found that signaling process was a pivotal condition to ensure the ovule develop normally in *Arabidopsis*, in which signal transduction was damaged in the female semisterile mutant [41]. *TET* genes control the cell proliferation, cell differentiation and cell identity, moreover, which also expressed in reproductive tissues and affected FG development in *Arabidopsis* [42, 43]. *RTNLB* contributes significantly to endoplasmic reticulum (ER) modeling, which plays an important role in linking the cortical ER to the plasma membrane and intracellular trafficking [44, 45]. In this study, the expression level of *TET18*, *RTNLB8* was markedly higher in FER than in STE. We calculated that the low expression of *TET18* and *RTNLB8* in STE could damage the cell-cell communication and the tubulation of the ER, and affect the intracellular transport and the cell structure. Kayes and Clark found that in *Arabidopsis* *CLV* genes can control vegetative and reproductive growth [46], and regulate the number of carpels and seeds [47]. Our results showed that *CLV1* was high expressed in FER, especially during the FNM1 period, while low expressed during the whole development period in the STE ovule. We predicted that the low expression of *CLV1* could result in cell signaling impaired and the failure of division and differentiation, finally made the ovule development damaged (Fig. 7).

Previous studies found that several regulatory genes of mitosis and apoptosis have been identified. For example, *CYCB* high expressed during the cell cycle and degrade in mitosis, thus controlling enter and retreat from M phase in *Nicotiana sylvestris* [48]. *TUBA* participates in mitosis, chromosome movement, and nuclear movement, mutations destabilize spindle microtubules in *Aspergillus nidulans* [49]. Several studies reported that *DHFR* high expressed at the G1/S of the cell cycle, which increased as cells undergo DNA synthesis and decrease as cells enter quiescence or terminally differentiate [50, 51]. In our investigation, with FNMM developing, *CYCB* and *TUBA* were significantly up-regulated in FER, *DHFR* also high expressed in FNM1 period, on the contrary, they all down-regulated in STE. It is suggested that mitosis maybe arrested at G2/M phase, DNA and some other biomolecules insufficient accumulation in G1/S, together with spindle formation was abnormal, which led free nuclear division accomplish difficulty in STE. Gao et al. found that the overexpression of *SOBIR1* leads to activation of cell death [52], which mutation can rescue the defects in the Golgi structure and premature shedding of floral organs, as well as participating in H₂O₂ regulation in *Arabidopsis* [53, 54]. Our results showed that *SOBIR1* is up-regulated in STE, especially in FNM3 stage, which might induce FG apoptosis and impact the vesicle transportation, also affect the accumulation of H₂O₂. Furthermore, in *Arabidopsis* *CPI* encodes enzymes which emerged as key in the regulation of PCD and helps cells avoid active oxygen damage as inhibited the H₂O₂ amplification [55]. We found that *CPI* nearly no expressed in STE, which may lead to PCD and the accumulation of active oxygen. Taken together, we conclude that the differences of mitosis and apoptosis between two lines may play an important role in ovule development to deal with FNM (Fig. 7).

The accumulation of H₂O₂ led to marked increase in lignin content and accelerate cell death [56]. *PER* regulates a variety of oxidation-reduction reactions to scavenge reactive oxygen, avoiding oxidative stress and participate in resistance against environmental stresses [57]. To resist the abiotic stresses, callose was formed to defense the organism and accelerate the formation of cell wall [58], Rosellini et al. found that callose deposition also existed in female sterile line of *Medicago sativa* L [5]. *CALS3* is responsible for the synthesis of callose in the temporary callose wall of the microspores and is essential for exine formation during microsporogenesis in *Arabidopsis* [59]. We found that *PER* was high expressed in FER, low expressed in the STE ovule. The expression level of *CALS* was complete opposite to *PER*. Thus, we proposed that the STE ovules might experience oxidative stress due to the incoordination active oxygen scavenging and H₂O₂ over-accumulation, result in a marked increase in lignin content and generated a series of nonfunctional protein. Moreover, the analysis of the SMRT and NGS

showed that the *UBQ* gene high expressed in FER and low expressed in STE, showing that protein degradation was not thorough in STE. Concurrently, the formation of callose and cell wall was activity in the STE ovules, those metabolic disorders finally led ovule abortion.

In addition, SMRT sequencing complements the weakness of NGS and provides advantages for discovery of AS. Previous studies show that the rate of AS is especially lower in plants [60], and there is a strong link between AS and evolution [61]. Loraine et al. revealed that AS participants in the reproduction development of *Arabidopsis* [62]. AS has been reported to regulate the formation of male and female gametogenesis in *Arabidopsis* [63], which also led to the male sterile in rice [23]. In this study, we have detected seven AS modes in the ovule of *P. tabuliformis*, there are more AS events in the FER than STE ovule (Fig. 4B, 4D). However, the number of AS events was significantly lower than the other species [22], which demonstrated that *P. tabuliformis* is more original. In addition, AS has been searched around the DEGs with opposite expression patterns during the FNM process between FER and STE, which might led to ovule abortion in *P. tabuliformis*. We found that isoforms of *AGPL1*, *bHLH66*, and *TUBA* were different in two lines (Fig. 4E), which influenced the starch metabolism and mitosis. We calculated that the quantity difference of AS events in varying lines, especially the different isoforms of *AGPL1*, *bHLH66*, and *TUBA* between two lines might lead to the development of ovules abnormal in STE.

Developing SSR markers is another merit of the SMRT sequencing. SSR marker has been used to determining protein function, genetic development, and gene expression regulation [24]. For example, SSRs marker has been used to identify the position of the gene caused female sterile in wheat and soybean [25, 26]. Several studies about gymnosperms already focused on SSRs, such as *Taxaceae*, *Pinus massoniana* and *Ginkgo biloba* L. [64-66]. Similarly to the other gymnosperms [66], excluded the mono-nucleotide repeats, di-nucleotide repeats were the most abundant type, followed by tri-nucleotide repeats. the most abundant motif was AT/AT; AAG/CTT, AGC/CTG, and AAT/ATT were the main repeat motifs of tri-nucleotide motifs in both two lines (Fig. 5). Undoubtedly, the transcriptome sequencing based on SMRT and NGS obtained more information than which only rely on a single database. The potential SSRs identified in this research provided productive resources for further investigation of SSRs in *P. tabuliformis* and established foundation to the development of molecular markers of gymnosperms, which would be valuable in helping with molecular study in this unique species in the future.

Conclusions

In this study, we analyzed transcriptome features based on the NGS and SMRT sequencing of the FER and STE ovules, which develop in different periods in *P. tabuliformis*. Our findings showed that, compared with NGS, SMRT sequencing overcome the limit of the length of reads, generated more transcriptome information. The DEGs associated with the carbohydrate metabolism, signal transduction, mitosis and apoptosis were determined, suggesting the energy metabolism, regulation of cell cycle and the accumulation of reactive oxygen and lignin were abnormal of STE ovules. Moreover, AS might play an important role in the growth and development of ovules which may be affected the fertility in *P. tabuliformis*. In addition, we also provided valuable resources for SSR markers development. These findings may provide new insights into ovule abortion in plants, especially highlight our understanding of the molecular mechanisms involved in the FNM process and ovule development in the conifers.

Methods

Plant materials and RNA sample preparation

Our research group FER found that the FER ovule and STE ovule were genetically closely related by DNA marker technique. Therefore, the FER and STE ovules were suitable to explore the mechanism of ovule abortion in conifer. Then, we selected three FER trees and STE trees were in the *P. tabuliformis* Seed Garden, Xingcheng, Liaoning Province, China. Material collection for this research was permitted by the Forestry Administration of Xingcheng. Cones of similar size were selected from the trees growing in common habitats, gathered in the early February (early period of FNM process), early March (middle period of FNM process) and early April (late period of FNM process), samples were collected from the middle of the crown of each tree. The

ovules were removed from the placenta of scales under a dissection microscope, each sample including hundreds of ovules from five cones. These samples were placed into liquid nitrogen overnight, then stored at -80°C .

Total RNA was extracted by grinding tissue using Plant RNA Assistant Kit (Kebaiao, Beijing, China) according to the manufacturer's instruction. The integrity of the RNA was quantified by Agilent 2100 Bioanalyzer and agarose gel electrophoresis. The purity and concentration of the RNA were determined with the Nanodrop 2000 (Thermo Specific, USA). The total RNA was prepared for two experiments: 1. the RNA-seq samples of FER and STE stages were used to construct eighteen libraries (FER-FNM1, FER-FNM2, FER-FNM3, STE-FNM1, STE-FNM2, and STE-FNM3; each sample had three replicates) and were subjected to 2*150 paired-end RNA-seq using Illumina HiSeqTM 4000; 2. total RNAs with an RNA integrity number (RIN) value larger than 8 were mixed equally from three development stages (FNM1, FNM2, FNM3) of FER and STE for a single molecular real-time (SMRT) sequencing using the PacBio RS II platform.

Analysis of PacBio SMRT

The SMRT Link v5.0.1 pipeline [67] supported by Pacific Biosciences was used to classify and cluster the raw sequencing reads of cDNA libraries into transcript consensus. In short, CCS (circular consensus sequence) reads were extracted and classified into short reads, full-length chimeras reads, non-full-length reads, and full-length non-chimeric (FLNC) based on cDNA primers and polyA tail signal. Then the FLNC reads were clustered by Iterative Clustering for Error Correction (ICE) software to generate the cluster consensus isoforms. The final transcriptome isoform sequences were filtered with software CD-HIT-v4.6.7 to remove the redundant sequences using a threshold of 0.99 identities.

Functional Annotation

Gene Ontology (GO) annotation was analyzed by Blast2GO software [68] with Nr annotation results of isoforms, and the functional classification of isoforms was performed using WEGO software [69]. The KEGG database was used to explore the genes involved in the biological pathways. The pathway was defined as a significantly enriched pathway for genes, if the p-value < 0.05.

Analysis of Differentially Expressed Genes (DEGs)

The reads per kilobase of exon model per million mapped reads (RPKM) was generally used to estimate the gene expression level [70]. The DEGs were identified based on the Audic and Claverie's method [71]. DEGs were identified by IDEG6 software (http://telethon.bio.unipd.it/bioinfo/IDEG6_form/). All statistical tests for this research were calibrated for multiple testing with the Benjamini-Hochberg false discovery rate (FDR). The FDR was less than 0.01, and the absolute value of log2 Ratio more than 1 was used to confirm significant differences in gene expression.

Alternative splicing (AS) detection

To analyze AS events of transcript isoforms, depending on k-mer similarity we used the COding GENome reconstruction Tool (Cogent) [22] to divide transcripts into gene families, De Bruijn graph methods were also used to reestablish each family into a coding reference genome. Furthermore, the transcript isoforms analyze of the AS events was by SUPPA [72] tool.

Simple sequence repeat (SSR) marker prediction

The MISA (<http://pgrc.ipk-gatersleben.de/misa/>) was used to discover the simple sequence repeats (SSRs) in the whole transcriptome. The minimum number of repeat units as follows: six for di-nucleotides, five for tri-nucleotides, four for tetra-, penta- and hexa-nucleotides. In addition, the shortest distance between two SSRs was 100bp, or they will be seen as one SSR.

Real-Time Quantitative PCR Analysis

Total RNA was isolated using Plant RNA Assistant Kit (Kebaiao, Beijing, China). Reverse transcription into cDNA was performed with SuperReal PreMix Plus (TIANGEN Biotech, Beijing, China). All of the experiments were performed following the protocols included with the kits. Primers were designed using Primer Premier 6. The qRT-PCR was performed using a MiniOpticon Two-Color Real-Time PCR Detection System (BIO-RAD, Hercules, CA, USA).

Abbreviations

FER: Female fertile line; STE: Female sterile line; FNM: Free nuclear mitosis; SMRT: Single molecule real-time; NGS: Next-generation sequencing; FLNC: Full-length non-chimeric; DEGs: Differentially expressed genes; AS: Alternative splicing; SSRs: Simple sequence repeats; FNMM: Free nuclear mitosis of megagametophyte; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; RPKM: Reads Per Kilobase per Million mapped reads; FDR: False discovery rate; qRT-PCR: Quantitative real-time PCR.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed in this study are included in this paper or its supplementary information files. The raw sequence data reported in this research have been deposited in the Genome Sequence Archive in BIG Data Center, Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under accession numbers , which are publicly accessible at <https://bigd.big.ac.cn/gsa>.

Competing interests

All the authors agreed on the contents of the paper and post no conflicting interest.

Funding

This work was supported by the Fundamental Research Funds for the Central Universities (BLX201617), and the National Natural Science Foundation of China (31700250, 31870571). The funding bodies didn't play a role in the design of the study and collection, analysis and interpretation of data and in writing the manuscript.

Authors' contributions

Z.G., C.Z. and Y.Z. designed the research; Z.G. and R.H. performed the research; Z.G. and R.H. analyzed data; Z.G., M.ZH., Q.Y. and M.Z. read and edited the paper, Z.G. wrote the article. All the authors have read and approved the final manuscript.

Acknowledgements

We thank Guangzhou Genedenovo Biotechnology Co., Ltd and the OmicShare forum, a free online platform for data analysis (www.omicshare.com/tools) in assistance with sequencing and bioinformatics analysis.

References

1. GuoA,ZhengC:Female gametophyte development.Journal of Plant Biology2013,56:345–356.

2. HuQ,GaoS,LiF:Advances of research on plant female sterility.Journal of Beijing Forestry University2004,26(1):87–91.
3. ChenL,ZhangJ,LiH,NiuJ,XueH,LiuB,WangQ,LuoX,ZhangF,ZhaoDetal:Transcriptomic Analysis Reveals Candidate Genes for Female Sterility in Pomegranate FERowers.Frontiers in plant science2017,8.
4. AwasthiA,PaulP,KumarS,VermaSK,PrasadR,DhaliwalHS:Abnormal endosperm development causes female sterility in rice insertional mutant OsAPC6.Plant Science2012,183:167–174.
5. RoselliniD,FerrantiF,BaroneP,VeronesF:Expression of female sterility in alfalfa (Medicago sativa L.).Sex Plant Reprod2003,15:271–279.
6. WiniarczykK,KosmalaA:Development of the female gametophyte in the sterile ecotype of the bolting Allium sativum L.Scientia Horticulturae2009,121:353–360.
7. ClarkSE,WilliamsRW,MeyerowitzEM:The CLAVATA1 Gene Encodes a Putative Receptor Kinase That Controls Shoot and floral Meristem Size in Arabidopsis.Cell1997,89:575–585.
8. ChalivendraS,HuberSC,SachsMM,RhoadsD:Sucrose Synthase.Plant Signal Behavior2007,2(1).
9. Kong. J,Li.Z,Tan.Y-P,Wan. C-X,Li.S-Q,ZhuY-G:Different gene expression patterns of sucrose-starch metabolism during pollen maturation in cytoplasmic male-sterile and male-fertile lines of rice.Physiologia plantarum2007,130:136–147.
10. Durán-FigueroaN,Vielle-CalzadaJ-P:ARGONAUTE9-dependent silencing of transposable elements in pericentromeric regions of Arabidopsis.Plant signaling & behavior2010,5(11):1476–1479.
11. HaveckerER,WallbridgeLM,HardcastleTJ,BushMS,KellyKA,DunnRM,SchwachF,DoonanJH,BaulcombeDC:The Arabidopsis RNA-directed DNA methylation argonautes functionally diverge based on their expression and interaction with target loci.The Plant cell2010,22(2):321–334.
12. Huang J, Wijeratne A, Tang C, Zhang T, Fenelon R, Owen H, Zhao D:Ectopic expression of TAPETUM DETERMINANT1 affects ovule development in Arabidopsis.Journal of Experimental Botany2016,67(5):1311–1326.
13. KempeK,RubtsovaM,RieweD,GilsM:The production of male-sterile wheat plants through split barnase expression is promoted by the insertion of introns and FERexible peptide linkers.Transgenic research2013,22(6):1089–1105.
14. LiuX,MeiW,SoltisPS,SoltisDE,BarbazukWB:Detecting alternatively spliced transcript isoforms from single-molecule long-read sequences without a reference genome.Molecular ecology resources2017,17(6):1243–1256.
15. TilgnerH,GrubertF,SharonD,SnyderMP:Defining a personal, allele-specific, and single-molecule long-read transcriptome.Proceedings of the National Academy of Sciences2014,111(27):9869–9874.
16. ChenX,LiQ,LiY,QianJ,HanJ:Chloroplast genome of Aconitum barbatum var. puberulum (Ranunculaceae) derived from CCS reads using the PacBio RS platform.Frontiers in plant science2015,6.
17. WangB,TsengE,RegulskiM,ClarkTA,HonT,JiaoY,LuZ,OlsonA,SteinJC,WareD:Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing.Nature communications2016,7(1).
18. BlencoweBJ:Alternative Splicing: New Insights from Global Analyses.Cell2006,126(1):37–47.
19. JiaD,WangY,LiuY,HuJ,GuoY,GaoL,MaR:SMRT sequencing of full-length transcriptome of FERea beetle Agasicles hygrophila (Selman and Vogt).Scientific reports2018,8(1).
20. ZhuFY,ChenMX,YeNH,ShiL,MaKL,YangJF,CaoYY,ZhangY,YoshidaT,FernieARetal:Proteogenomic analysis reveals alternative splicing and tranSTEation as part of the abscisic acid response in Arabidopsis seedlings.The Plant Journal2017,91(3):518–533.
21. XuZ,PetersRJ,WeiratherJ,LuoH,LiaoB,ZhangX,ZhuY,JiA,ZhangB,HuSetal: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.The Plant Journal2015,82(6):951–961.
22. LiJ,Harata-LeeY,DentonMD,FengQ,RathjenJR,QuZ,AdelsonDL:Long read reference genome-free reconstruction of a full-length transcriptome from Astragalus membranaceus reveals transcript variants involved in bioactive compound biosynthesis.Cell discovery2017,3:17031.
23. SunL,ZhangL,ZhangP,YangZ,ZhanX,ShenX,ZhangZ,HuX,XuanD,WuWetal:Characterization and Gene Mapping of an Open Hul Male Sterile Mutant ohms1 Caused by Alternative Splicing in Rice.Chinese Journal of Rice Science2015,29(5):457–

24. LawsonMJ,ZhangL:Distinct patterns of SSR distribution in the Arabidopsis thaliana and rice genomes.Genome biology2006,7(2):R14.
25. DouB,HouB,XuH,LouX,ChiX,YangJ,WangF,NiZ,SunQ:Efficient mapping of a female sterile gene in wheat (Triticum aestivum L.).Genetics Research2009,91(5):337–343.
26. KatoKK:Molecular Mapping of the Male-Sterile, Female-Sterile Mutant Gene (st8) in Soybean.Journal of Heredity2003,94(5):425–428.
27. ZhangM,ZhangW,GongZ,ZhengC:Morphologic and anatomical observations in the process of ovulate strobilus generation and development in Pinus tabuliformis.Journal of Beijing Forestry University2017,39(6):1–12.
28. ChenK,AbbottRJ,MilneRI,TianXM,LiuJ:Phylogeography of Pinus tabulaeformis Carr. (Pinaceae), a dominant species of coniferous forest in northern China.Molecular ecology2008,17(19):4276–4288.
29. ZhangM,ZhengC-x:Archegonium and fertilization in Coniferopsida.Trees2016,30(1):75–86.
30. LiF,ZhengC:A Preliminary Study on Ovule Abortion During Later Stage of Development in Pinus tabulaeformis.Journal of Beijing Forestry University1990,12(3):68–73.
31. YaoY,HanR,GongZ,ZhengC,ZhaoY:RNA-seq Analysis Reveals Gene Expression Profiling of Female Fertile and Sterile Ovules of PinusTabulaeformis Carr. during Free Nuclear Mitosis of the Female Gametophyte.International journal of molecular sciences2018,19(8).
32. SolomonM,BelenghiB,DelledonneM,MenachemE,LevineA:The Involvement of Cysteine Proteases and Protease Inhibitor Genes in the Regulation of Programmed Cell Death in Plants.The Plant cell1999,11:431–443.
33. Ner-GaonH,HalachmiR,Savaldi-GoldsteinS,RubinE,OphirR,FERuhrR:Intron retention is a major phenomenon in alternative splicing in Arabidopsis.The Plant Journal2004,39(6):877–885.
34. WangH,LiuR,WangJ,WangP,ShenY,LiuG:The Arabidopsis kinesin gene AtKin–1 plays a role in the nuclear division process during megagametogenesis.Plant cell reports2014,33(5):819–828.
35. Lo´pez-AlmansaJC,PannellJR,GillL:Female sterility in Ulmus minor (Ulmaceae): a hypothesis invo king the cost of sex in a clonal plant.American Journal of Botany2003,90(4):603–609.
36. AndersonSN,JohnsonCS,JonesDS,ConradLJ,GouX,RussellSD,SundaesanV:Transcriptomes of isolatedOryza sativagametes characterized by deep sequencing: evidence for distinct sex-dependent chromatin and epigenetic states before fertilization.The Plant Journal2013,76(5):729–741.
37. YangL,WuY,YuM,MaoB,ZhaoB,WangJ:Genome-wide transcriptome analysis of female-sterile rice ovule shed light on its abortive mechanism.Planta2016,244(5):1011–1028.
38. ItoA,HayamaH,KashimuraY:Sugar metabolism in buds during FERower bud formation: a comparison of two Japanese pear [Pyrus pyrifolia (Burm.) Nak.] cultivars possessing different FERowering habits.Scientia Horticulturae2002,96:163–175.
39. BaezM,CabreraR,Guixev,BabulJ:Unfolding Pathway of the Dimeric and Tetrameric Forms of Phosphofructokinase–2 from Escherichia coli.Biochemistry2007,46:6141–6148.
40. KotlarzD,BuchH:Regulatory Properties of Phosphofructokinase 2 from Escherichia coli.European journal of biochemistry1981,117:569–574.
41. HuckN:The Arabidopsis mutant feronia disrupts the female gametophytic control of pollen tube reception.Development2003,130(10):2149–2159.
42. ReimannR,KostB,DettmerJ:TETRASPANINs in Plants.Frontier in Plant Science2017,8:545.
43. WangF,MutoA,Van deVeldeJ,NeytP,HimanenK,VandepoeleK,VanLijsebettensM:Functional Analysis of the Arabidopsis TETRASPANIN Gene Family in Plant Growth and Development.Plant physiology2015,169(3):2200–2214.
44. KriechbaumerV,BotchwaySW,STEadeSE,KnoxK,FrigerioL,OparkaK,HawesC:Reticulomics: Protein-Protein Interaction Studies with Two Plasmodesmata-Localized Reticulon Family Proteins Identify Binding Partners Enriched at Plasmodesmata, Endoplasmic Reticulum, and the Plasma Membrane.Plant physiology2015,169(3):1933–1945.

45. TolleyN,SparkesI,CraddockCP,EastmondPJ,RunionsJ,HawesC,FrigerioL:Transmembrane domain length is responsible for the ability of a plant reticulon to shape endoplasmic reticulum tubules in vivo.The Plant Journal2010,64(3):411–418.
46. KayesJM,ClarkSE:CLAVATA2, a regulator of meristem and organ development in Arabidopsis.Development1998,125:3843–3851.
47. FerrándizC,PelazS,YanofskyMF:Control of carpel and fruit development in Arabidopsis.Annual Review of Biochemistry1999,68:321–354.
48. TrehinC,Ahnl-O,PerennesC,CouteauF,LalanneE,BergouniouxC:Cloning of upstream sequences responsible for cell cycle regulation of the *Nicotiana sylvestris* CycB1;1 gene.Plant molecular biology1997,35:667–672.
49. GambinoJ,BergenLG,MorrisNR:Effects of mitotic and tubulin mutations on microtubule architecture in actively growing protoplasts of *Aspergillus nidulans*.The Journal of Cell Biology1984,99:830–838.
50. GorelovaV,DeLepeleireJ,VanDaeleJ,PluimD,MeiC,CuyppersA,LerouxO,RebeilleF,SchellensJHM,BlancquaertDetal:Dihydrofolate Reductase/Thymidylate Synthase Fine-Tunes the Folate Status and Controls Redox Homeostasis in Plants.The Plant cell2017,29(11):2831–2853.
51. STEanskyJE,FarnhamPJ:Transcriptional regulation of the dihydrofolate reductase gene.Bioessays1996,18(1):55–62.
52. LiuJ,DingP,SunT,NittaY,DongO,HuangX,YangW,LiX,BotellaJR,ZhangY:Heterotrimeric G Proteins Serve as a Converging Point in Plant Defense Signaling Activated by Multiple Receptor-Like Kinases.Plant physiology2013,161(4):2146–2158.
53. GaoM,WangX,WangD,XuF,DingX,ZhangZ,BiD,ChengYT,ChenS,LiXetal:Regulation of cell death and innate immunity by two receptor-like kinases in Arabidopsis.Cell host & microbe2009,6(1):34–44.
54. LeSTEieME,LewisMW,YounJY,DanielsMJ,LiljegrenSJ:The EVERSHED receptor-like kinase modulates floral organ shedding in Arabidopsis.Development2010,137(3):467–476.
55. TiwariBS,BelenghiB,LevineA:Oxidative stress increased respiration and generation of reactive oxygen species, resulting in ATP depletion, opening of mitochondrial permeability transition, and programmed cell death.Plant physiology2002,128(4):1271–1281.
56. MilosevicN,STEusarenkoAJ:Active oxygen metabolism and lignification in the hypersensitive response in bean.Physiological and Molecular Plant Pathology1996,49:143–158.
57. KawanoT:Roles of the reactive oxygen species-generating peroxidase reactions in plant defense and growth induction.Plant cell reports2003,21(9):829–837.
58. LunaE,PastorV,RobertJ,FERorsV,Mauch-ManiB,TonJ:Callose deposition: a multifaceted plant defense response.Molecular Plant Microbe Interactions2011,24(2):183–193.
59. DongX,HongZ,SivaramakrishnanM,MahfouzM,VermaDP:Callose synthase (CalS5) is required for exine formation during microgametogenesis and for pollen viability in Arabidopsis.The Plant Journal2005,42(3):315–328.
60. BrettD,PospisilH,ValcárcelJ,ReichJ,BorkP:Alternative splicing and genome complexity.Nature Genetics2001,30(1):29–30.
61. GamazonER:Alternative Splicing and Genome Evolution.2016:1–6.
62. LoraineAE,McCormickS,EstradaA,PatelK,QinP:RNA-Seq of Arabidopsis Pollen Uncovers Novel Transcription and Alternative Splicing.Plant physiology2013,162(2):1092–1109.
63. ForsthoefelNR,KlagKA,McNicholSR,ArnoldCE,VernonCR,WoodWW,VernonDM:Arabidopsis PIRL6 Is Essential for Male and Female Gametogenesis and Is Regulated by Alternative Splicing.Plant physiology2018,178(3):1154–1169.
64. FengYH,YangZQ,WangJ,LuoQF,LiHG:Development and characterization of SSR markers from *Pinus massoniana* and their transferability to *P. elliottii*, *P. caribaea* and *P. yunnanensis*.Genetics and Molecular Research2014,13(1):1508–1513.
65. FuWEI,DuZ-X,RenHE,BuahomN,LiuH-W,HuangXIN:Polymorphic microsatellite markers in *Taxus chinensis* var. *mairei* (Taxaceae).Journal of Genetics2013,93(S2):81–84.
66. HanS,WuZ,JinY,YangW,ShiH:RNA-Seq analysis for transcriptome assembly, gene identification, and SSR mining in ginkgo (*Ginkgo biloba* L.).Tree Genetics & Genomes2015,11:37.
67. GordonSP,TsengE,SalamovaA,ZhangJ,MengX,ZhaoZ,KangD,UnderwoodJ,GrigorievIV,FigueroaMetal:Widespread Polycistronic Transcripts in Fungi Revealed by Single-Molecule mRNA Sequencing.PloS one2015,10(7):1–15.

68. ConesaA,GötzS,García-GómezJM,TerolJ,TalónM,RoblesM:Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.Bioinformatics2005,21(18):3675–3676.

69. YeJ,FangL,ZhengH,ZhangY,ChenJ,ZhangZ,WangJ,LiS,LiR,BolundLetal:WEGO: a web tool for plotting GO annotations.Nucleic Acids Research2006,34(Web Server):W293-W297.

70. AudicS,ClaverieJM:The significance of digital gene expression profiles.Genome Res1997,7(10):986–995.

71. MortazaviA,WilliamsBA,McCueK,SchaefferL,WoldB:Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods2008,5(7):621–628.

72. AlamancosGP,PagèsA,TrincadoJL,BelloraN,EyrasE:Leveraging transcript quantification for fast computation of alternative splicing profiles.RNA2015,21(9):1521–1531.

Table

Table 1. Summary of SSRs identified in *P. tabuliformis* transcriptome of FER and STE ovule

Searching item	Numbers	
	FER	STE
Total number of sequences examined	40828	44195
Total size of examined sequences (bp)	112127427	98231600
Total number of identified SSRs	5530	5096
Number of SSR containing sequences	4276	3948
Number of sequences containing more than 1 SSR	868	796
Number of SSRs present in compound formation	475	487
Di-nucleotide	2516	2485
Tri-nucleotide	2132	1846
Tetra-nucleotide	463	407
Penta-nucleotide	134	130
Hexa-nucleotide	285	228

Figures

Figure 1

Flowchart of the experimental design and analysis for SMRT sequencing and RNA sequencing. Schematic graph of free nuclear mitosis (FNM) during ovule development in female fertile lines (FER) and sterile lines (STE) of *P. tabulaeformis* collected for SMRT sequencing and Illumina sequencing. Combine PacBio and RNA-seq were used for a series of analyses, including alternative splicing (AS), DEG analysis, and simple sequence repeats (SSRs). Finally, revealing the transcriptome differences between FER and STE.

Figure 2

Summary of SMRT sequencing proportions of four kinds of the circular consensus sequence (CCS), which classified into full-length chimeras reads, full-length non-chimeric reads (FLNC), non-full-length reads, and short reads (A); length distribution of FLNC, FER (B), STE (C); DEGs with opposite expression patterns during the FNM process in STE and FER, DEGs up-regulated in FER and down-regulated in STE (D); DEGs up-regulated in STE and down-regulated in FER (E); comparison of transcript length distribution from different sequencing platforms (F); GO classification of DEGs up-regulated in FER also down-regulated in STE (G); GO classification of DEGs up-regulated in STE also down-regulated in FER (H).

Figure 3

Heat map diagram of expression levels of DEGs expression level exist obvious differences in different lines. Data for the gene expression level were normalized to Z-score.

Figure 4

Comparison of different alternative splicing(AS) among FER and STE. Schematic diagram of Cogent assemble (A); visualization of seven AS modes (B); distribution of isoform numbers in FER and STE (C); different types of AS events in FER and STE (D); the model graph shows AGPL1/bHLH66/TUBA generating different transcript isoforms of RI detected in two tissues (E).

Figure 5

Comparison of different simple sequence repeat (SSR) among FER and STE. The distribution of SSRs based on the number of repeat units in FER (A) and STE (B); frequency distribution of SSRs based on motif types in FER (C) and STE (D).

Figure 6

Candidate DEGs expression levels revealed by qRT-PCR. Sucrose synthase (SUS); peroxidase 12 (PER12); ATP-dependent 6-phosphofructokinase 2 (PFK2); clavata 1-like protein (CLV1); leucine-rich repeat receptor-like serine/threonine/tyrosine-protein kinase SOBIR1 (SOBIR1); polyubiquitin-like 10 (UBQ10); tubulin alpha chain (TUBA); ADP-glucose pyrophosphorylase large subunit 1 (AGPL1); tetraspanin-18 (TET18)

Figure 7

Speculative network model revealing the ovule abortion in the female-sterile line of *Pinus tabulaeformis* Carr. SUS: Sucrose synthase; AGPL1: ADP-glucose pyrophosphorylase large subunit 1; PFK2: ATP-dependent 6-phosphofructokinase 2; TET18: Tetraspanin-18; NPF 3.1: NRT1/PTR family 3.1; RTNLB8: Reticulon-like protein B8; CLV1: Clavata 1-like protein; CYCB: Cyclin B; TUBA: Tubulin alpha chain; DHFR: Dihydrofolate reductase; SOBIR1: Leucine-rich repeat receptor-like serine/threonine/tyrosine-protein kinase SOBIR1; CPI: Cysteine proteinase inhibitor; UBQ: Polyubiquitin-like; PER: Peroxidase; exgA: Glucan 1,3-beta-glucosidase A; CALS3: Callose synthase 3; bHLH: Basic helix-loop-helix proteins.

Supplementary Files

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

- [FigS1.pdf](#)
- [FigS2.pdf](#)
- [FigS3.pdf](#)
- [FigS4.pdf](#)
- [SupplementaryTablesS1S9.xlsx](#)