De novo assembly of expressed transcripts and analysis of pistil to identify genes involved in early stage of pollination in *Liriodendron chinense*

Ming Li[1], Kun Wang[2], Rebecca Njeri Damaris[1,3], and Pingfang Yang[1]*

1 Key Laboratory of Plant Germplasm Enhancement and Speciality Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China;
2 College of Life Science, Wuhan University, Wuhan, China
3 University of Chinese Academy of Sciences, Beijing 100039, China

Corresponding authors
* Pingfang Yang, Key Laboratory of Plant Germplasm Enhancement and Speciality Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China. Tel: +86-27-8751-0956, Fax: +86-27-8751-0956, E-mail: yangpf@wbgcas.cn

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

Abstract

Plant sexual reproduction is a complicated and a key biological process with profuse interactions between pollen and pistil. This process determines whether fertilization will be successful or not and thus affect the seed setting. To explore the reason why *L. chinense* has a low seed setting ratio, transcriptome analysis on pistils of *L. chinense* during pollination were conducted. After analyzing the sequencing data, 206,858 unigenes with an average length of 646 bp were generated using the assembled transcripts. Among total unigenes, 3844 genes which expression fold change during early stage of pollination was higher or lower than 10 were selected as significant differentially expressed genes. 54 differentially expressed genes involved in sexual reproduction processes including the regulation of pollen tube growth process and double fertilization process might be partially causing the low seed setting in *L. chinense*. These results indicated that the barrier between pollen tube and pistil might be the reason why *L. chinense* have low seed setting. This study might be helpful to understand why *L. chinense* has such a low seed setting ratio.

Keywords

Liriodendron chinense, transcriptome, pistil, sexual reproduction

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

1 Introduction

Liriodendron genus consists of two arborescent species which includes *Liriodendron chinense* (Hemsl.) Sarg. and *Liriodendron tulipifera* L. (Hereinafter referred to as *L. chinense* and *L. tulipifera* respectively). *L. chinense* commonly known as mandarin jacket tree due to its Qing dynasty mandarin shaped leaves. While *L. tulipifera* commonly known as yellow-poplar or tulip tree due to its flower shaped like tulip and its wood structure and density resemble true poplar (Liang et al. 2011). Besides its peculiar leave and flower structures, *L. chinense* is also an economically valuable tree with dense wood. Like other members of magnoliaceae, trees in the genus *Liriodendron* were once widely distributed in northern hemisphere (Latham and Ricklefs 1993). However, only two extant species (*L. chinense* and *L. tulipifera*) in the genus *Liriodendron* that survived in the last Ice Age and constitute a well-known example of an East/North American vicariant species pairs after long geological time (Hao et al. 1995). Even though they survived the last Ice Age, they still face harsh environmental conditions which endangers their existence, especially *L. chinense*, which has been regarded as a rare and endangered species because its small, insular and sparse populations (Hao et al. 1995). In 1990's, *L. chinense* was classified as endangered or near-threatened species in the Red List of Endangered Plant in China and IUCN Red List of Threatened Species respectively (Wang and Xie 1992).

As ideal plants for research on mating system, systemic evolution and population genetics, there has been research studies about development of EST-SSR markers and population genetic analysis (Xu et al. 2010; Yang et al. 2012; Yang et al. 2011; Yao et al. 2007; Zhu and Li 2010). As one of the only two surviving species in genus *Liriodendron*, *L. chinense* has been used extensively as key specimen species in plant evolution studies (Cai et al. 2006; Parks and Wendel 1990; Wen 1999). As an

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

endangered tree with low seed setting percentage, *L. chinense* has received more attention in research field in China. It was reported that the seed setting ratio were 0.45%-14.72% (Fang et al. 1994). There have been numerous researchers work on this species for years in past two decades, such as examining the development of female gametophyte (Qin and Li 1996), pollen fertility (Zhou and Fan 1994), pollination (Huang and Guo 2002; Zhou and Fan 1999), flower or seed predation and genetics (Huang et al. 1999; Huang et al. 1998). Based on previous research, the ability of pollen were normal, while embryo sac exist abortive situation in different populations (Qin and Li 1996). Furthermore, *L. chinense* is tend to be cross-pollinated by insects but without faithful pollinator (Huang et al. 1999). Unfortunately, there is no conclusive agreement on the reasons why *L. chinense* has such a low seed setting ratio.

Based on the previously research, *L. chinense* has been classified in the magnoliaceae family and has the lowest chromosome number ($2n=2x=38$) (Meng et al. 2006). It is a dicot and has an estimated haploid genome size of 1,802 Mbp (Liang et al. 2007b). In the recent past, *L. tulipifera* in genus *Liriodendron* was sequenced using 454 pyrosequencing technology, and a high-quality EST database which includes 6520 unigenes was constructed (Liang et al. 2007a). Based on next-generation sequencing technology, RNA sequencing for transcriptome methods is suitable for specific species with genome not fully sequenced (Huang et al. 2012a; Li et al. 2012b; Liu et al. 2013; Wang et al. 2013b; Ward et al. 2012). In *L. chinense,* miRNAs in somatic embryos of *L. chinense* were analyzed by Solexa sequencing technology, 232 conserved miRNAs including 177 species-specific miRNAs were detected (Li et al. 2012c). A transcriptome analysis of petals and leaves showed that 3386 genes differentially expressed between petals and leaves (Yang et al. 2014). As high throughput omics technique, transcriptome is an effective and economical way to dig more candidate genes involved in biological process including reproduction process. However, there is hardly any omics report related to reproduction in *L. chinense.*

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

For better understand the reason why *L. chinense* has such a low reproductive efficiency from RNA and protein level, high throughput sequencing technology was applied in our lab. Initially, deep sequencing technology was applied to discover miRNAs in *L. chinense* flowers at different development stages. A total of 496 conserved miRNAs were identified and ten percent targets genes of conserved miRNAs were highly associated with the reproductive process (Wang et al. 2012a). Secondly, proteomics technology was applied to find out which proteins are involved in pollination. Proteomic analysis revealed that 493 proteins changed their expressions after pollination and 66 of them are involved in reproductive process. Among the reproductive process related proteins, protein disulfide-isomerase A6 and four embryo-defective proteins were highly associated with low reproductive efficiency (Li et al. 2014). In this present study, illumina paired-end solexa sequencing platform was employed to characterize different tissues *de novo* transcriptome in *L. chinense*, find out the predominantly expressed transcripts in pistil, and identify differentially expressed transcripts in pistil after pollination. A large set of *L. chinense* transcript sequences were obtained to showing the different expressed genes involved in *L. chinense* during pollination. This comprehensive transcriptome data set for *L. chinense*, which containing 206,858 unigenes, representing extensive putative functions will provide public information database for understanding the molecular mechanisms of reproduction in *L. chinense*.

2 Materials and methods

2.1 Tissue source

The *L. chinense* plants which were 15-20 years old were grown in Wuhan Botanical Garden, Chinese Academy of Sciences (N30°32′50.83″, E114°25′1.80″). Pollen and pistil which are from five individual tree were collected and pooled as previously described (Li et al. 2014). In order to keep the pistils in same environment during pollination, tree branches with flower buds about to open were cut from the tree on 28th April, 2013, and then they were cultured with half-strength Hoagland's

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

nutrient solution in growth chamber until pollination. The growth condition were set as: 14 h light (400-800 μmol m^{-2} s^{-1}) at 26 °C, 10 h darkness at 20 °C, and the relative humidity was maintained at 60-70% (Hoagland and Arnon 1950; Zhou and Fan 1994). Mature pollen grains were collected from the opening flowers and stored in -80 °C. For artificial pollination, the flower buds with half-open petals were chosen and the androecium was emasculated at night before pollination. After androecium emasculating, the un-pollinated pistils were cut from the flowers, and artificial pollination was done as follow: Mature pollen grains were harvested from open flowers and immediately smeared on the pistils of emasculated flowers using a soft brush immediately. After 30 minutes and 1h, the pollinated pistils were collected and stored in liquid nitrogen respectively, and then they were stored in -80 °C until use.

2.2 RNA extraction

Total RNAs were extracted from each tissue separately, according to the manufacturer's protocol for the TRIzol Reagent (Invitrogen, Shanghai, China). The extracted RNAs were treated with RNase-free DNase I (Takara, Dalian, China) to remove the DNA residues. After quality and purity assessment by Nanodrop 2000 (Thremo Scientific, Delaware, USA) and agarose gel, the RNA integrity numbers (RIN) were assessed by the Agilent 2100 Bioanalyzer (Agilent Technologies, California, USA). After that, approximately 500 μg of total RNA of each tissue was obtained. High quality mRNA of each tissue was purified using Illumina TruSeq™ RNA Sample Preparation Kit (Illumina, California, USA) according to the manufacturer's protocol.

2.3 cDNA library construction and Illumina sequencing

Poly(A) mRNAs were purified from the total RNA using Oligo(dT) RNA purification. The purified poly(A) RNAs were then dissolved into fragmentation solution (Ambion, California, USA) for 2.5 min at 94 ℃. After mRNAs fragmenting, first strand cDNA were synthesized using random hexamer-primer and these short fragments were taken as templates. The second strand cDNA was generated using the

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen, California, USA). These cDNA fragments then subjected to end repair process and A tailing, and then ligation of the sequencing adapters. These products were then purified and enriched with PCR to construct the final cDNA library. Finally, the cDNA library was sequenced using Illumina HiSeq100PE (Illumina, California, USA).

2.4 Data filtering and *de novo* transcriptome assembly

By using Fast QC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) software, raw reads generated by Illumina HiSeq pair-end 2X100 were filtered to obtain high quality clean reads by removing adaptor sequences, reads containing more than 10% Ns (N represents ambiguous bases), and low-quality reads defined as having more than 10% bases with Q value <20. All the following analyses were based on the clean reads. After data filtering, all the clean reads were assembled using software Trinity (Grabherr et al. 2011). Finally, the contigs that could not be extended on either end were defined as unique transcripts using Trinity.

2.5 Functional annotation and classification of the assembled unigenes

All of the assembled unigenes were searched in the public protein databases including Clusters of Orthologous Groups (COGs), Swiss-Prot protein database, NCBI nonredundant protein (Nr), Gene Ontology (GO), using the BLASTx analysis with a cut-off E value of 10^{-5} (Korf et al. 2003). Generally, all the assembled unigenes were searched against the Nr, COGs, and Swiss-Prot protein database using the BLAST algorithm with an E-value cut-off of 10^{-5}(Korf et al. 2003). Additionally, GO (Gene Ontology) terms were extracted from the best hits obtained from the BLASTx against the Nr using the Blast2GO program (Gotz et al. 2008). The best-aligning results were used to identify the sequence direction and to predict the coding regions. When the results from different databases conflicted, a priority order of Swiss-Prot, COG, KEGG and NR was followed. The ESTScan software was used for the analysis of unigenes that did not align to any of the above databases(Iseli et al. 1999). The CDS were selected from unigene sequences based on the blastx result, and

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

the CDS were also used to train ESTScan. The shortest CDS were at least 100 bp. In order to get the annotation information of assembled transcripts annotations, the BLAST2GO program was used to get GO annotations of unique assembled transcripts for describing biological processes, molecular functions, and cellular components(Conesa et al. 2005). Additionally, after getting GO annotations for each transcript, WEGO software was used to conduct GO functional classification for understanding the gene functions of the species from the macro level(Ye et al. 2006).

2.6 Identification of differentially expressed genes

For unigene expression analysis, the abundance of each unigene in each tissue sample was normalized and calculated by using the RPKM method (Mortazavi et al. 2008). Based on a previous study (Chen et al. 2011), the gene expression in none biological repeat sample will be in accordance with Poisson distribution, the Wald-Log test method is more powerful to detect the differentially expressed genes than other methods. For screening differentially expressed genes, the RPKM and Wald-log methods were used together. The threshold of p value in multiple tests was checked by manipulating the false discovery rate value (FDR≤0.01)(Benjamini and Hochberg 1995). In this study, the unigenes with ratios of RPKM between samples greater than 2 and FDR≤0.001 were considered to have significant differences in expression.

3 Results and discussion

3.1 RNA-seq and de novo transcriptome assembly

In order to identify the genes involved in early stage of pollination in *L. chinense*, determine the transcriptomes of *L. chinense* during pollination, cDNA libraries which constructed from RNA samples from un-pollinated pistil, pollinated pistil after 30min, and pollinated pistil after 1h were subjected to pair-end read sequencing with the Illumina platform. After deep sequencing of *L. chinense* transcriptome, 96367367 clean reads were harvested from a total of 97933908 raw reads, more than 19G data (Table 1). The sequencing raw data have been submitted to the Sequence Read

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

Archive (SRA) with the project ID: PRJNA277997 in NCBI database. These clean reads were assembled into 243,566 transcripts with an average length of 766 bp using Trinity software. Then 206,858 unigenes with an average length of 646 bp were generated using the assembled transcripts (Table 2).

Recently, one de novo transcriptome analysis of *L. chinense* petals and leaves was reported by a research group in Nanjing Agriculture University (Yang et al. 2014). Here we compare our data between the data produced by Nanjing Agriculture University (Hereinafter referred to Nanjing's data). First of all, the quality of two group's sequencing data was assessed using FastQC software. The quality of our sequencing data was higher than Nanjing's data in spite of different tissues between two groups (Sup Fig.1). Secondly, the de novo assembled results of two groups were compared. The assembled transcripts results are similar between our result and the results we assembled using data submitted by Nanjing group. However, there is difference between the assembled results which were assembled with data submitted by Nanjing group and the results in their paper (Sup table 1). Unfortunately, there is no detail description of this point in their paper.

3.2 Functional annotation and classification of the assembled unigenes

In order to understand the functions of these assembled unigenes, these unigenes were blasted to Clusters of Orthologous Groups (COGs), the Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), NCBI non-redundant protein (Nr), and Swiss-Prot protein database. Totally, 93,274 (45% of all unigenes) unigenes encoding products that have significant similarity to previously characterized proteins in one of the public databases mentioned above (Table 3).

With the GO classification, the 27,094 matched unigenes were classified into 3 functional categories: cellular components, molecular function, and biological process. In cellular components, these matched sequences were classified in 12 categories: the most represented cellular components were cell (13214, 31.54%) and cell part (13214, 31.54%) and the second group was organelle (8076, 19.28%). In molecular function,

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

the matched sequences were clustered into 13 functions including catalytic activity (16470, 50.51%), binding (12177, 37.35%), transporter activity (1920, 5.89%), and others. According to biological process, the matched sequences were classified into 23 categories containing metabolic process (14356, 25.24%), cellular process (13370, 23.51%), reproduction (992, 1.74%), and reproductive process (966, 1.70%) (Fig. 1).

Like other basal angiosperms (http://ancangio.uga.edu/content/aagp-home), there is scarcity of genome information in databases, and these basal angiosperms exhibit heterozygosity with complex and big sized genome. In this study, the rate of annotated unigenes (45%) was lower than the range of previous studies in other non-model species (92.65% in chili pepper, 92.09% in radish, 73.6% in blueberry, 58.01% in Chinese fir and 58% in safflower flowers), demonstrating its extreme lack of genome annotation information in Liriodendron (Huang et al. 2012b; Li et al. 2012a; Li et al. 2012d; Liu et al. 2013; Wang et al. 2013a). Recently, there has been a transcriptome analysis of *L. chinense* petals and leaves which showed that 74.6% unigenes of assembled unigenes were annotated (Yang et al. 2014). Comparing with their results, we found the total data size is similar between results in our group and Nanjing group, but the number of assembled transcripts was obviously different (Sup table 1). Furthermore, the transcripts were assembled using Nanjing's raw data which were deposited in NCBI, and the number of assembled transcripts was larger than Nanjing's report (Sup table 1). Although the annotated gene ratio was lower than Nanjing's results, the total number of annotated gene was larger than Nanjing's results (Sup Fig.2). However, it is not clear why there is such a large difference between these data. It might because of the difference cultivars or sequencing technology used in two groups.

3.3 Identification of differentially expressed genes during early stage of pollination

Totally, 19894 of 206,858 unigenes were identified as differentially expressed genes during early stage of pollination. In order to focus on genes which have close

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

relation with pollination, only 3844 genes which expression fold change during early stage of pollination was higher or lower than 10 were selected as significant differentially expressed genes (Sup Table 2). These 3844 significant differentially expressed genes were clustered into four groups. Cluster 1 was composed of 1251 genes that exhibited increased expression during pollination. Cluster 2 contained 624 genes that had a sustained decreased expression during pollination. Cluster 3 consisted of 960 genes that reached a peak expression level at 30 min after pollination and then decreased. Cluster 4 consisted of 1009 genes that decreased expression level at 30 min after pollination and then increased (Fig. 2; Sup Table 2).

3.4 Differentially expressed genes involved in early stage of sexual reproduction in *L. chinense*

These 3844 significant differentially expressed genes might involve in the regulation of pollination in *L. chinense*, however only 2440 genes have annotation in green plant database after blasting (Sup Table 2). Among them, 54 genes were annotated that they might involve in sexual reproduction process including pollen germination, regulation of pollen tube growth, double fertilization forming a zygote and endosperm, and others (Table 4). These differentially expressed genes might involve in pollination in *L. chinense.* Generally, there are two main kinds of genes might have close relationship with sexual reproduction. One is pollen tube growth related genes, among 54 sexual reproduction related genes, 25 genes were clustered into 4 clusters and annotated they might involve in pollen tube growth or the regulation of pollen tube growth processes (Table 4; Fig. 3A). Another one is double fertilization forming a zygote and endosperm related genes, among 54 sexual reproduction related genes, 9 genes were clustered into 3 clusters and annotated they might involve in double fertilization processes (Table 4; Fig. 3B).

3.5 Genes involved in the regulation of pollen tube growth processes

In cluster 1, four genes including S-locus lectin protein kinase family protein and DC1 domain-containing protein had a sustained decreased expression during

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

pollination. Proteins with S-locus glycoprotein domain are reported to be involved in self-incompatibility through inhibiting pollen tube growth from self-pollen (Sakamoto et al. 1998). In Brassicaceae, the *SLG* encodes a secretory glycoprotein which is synthesized in stigma predomaintly(Nasrallah et al. 1987). However, there is less research about molecular genetic analyses of *S* locus in *L. chinense.* In this study, two S-locus lectin protein kinase family proteins (one gene was in cluster 2) were identified have differential expression tendency. These results indicated that *L. chinense* might have similar self-incompatibility mechanism to recognize self-pollen. Another gene which homological gene in Arabidopsis is DC1 domain-containing protein (AT1G60420.1) was identified. C1 domains have been demonstrated to bind diacylglycerol and phorbol esters, and this domain is implicated in lipid signaling in mammals (Chen et al. 2008). In Arabidopsis, the *male sterile 1* pistils were pollinated with heterozygous pollen from single-locus T-DNA insertion of AT1G60420 line, and the results showed the defective pollen tube growth in the pistils(Qin et al. 2009). In consistence with the pollen growth in pistil of T-DNA insertion of AT1G60420 line in Arabidopsis, somewhat aberrantly shaped pollen tube growths were observed in *L. chinense* pistil (Li et al. 2014).

In cluster 3, five genes including gametophyte defective 1 and plantacyanin was no changed in pollinated pistil after 30 min and up regulated in pollinated pistil after 60 min. In previous study, it was proved that the lack expression of gametophyte defective 1 lead to aberrant development of embryo sac and reduced fitness of pollen. It also suggested that gametophyte defective 1 is essential for female gametogenesis and male fitness(Wang et al. 2012b). However, this gene just increased the expression in pollinated pistil after 60 min later, and this late expression in *L. chinense* pistil might have a similar function with lack expression in Arabidopsis. It was showed that the over expression of plantacyanin resulted in pollen tubes could not grow toward the style/ovary, making many turns around the papillar cell surface and ending their growth at the papillar cell tip (Chae and Lord 2011). We also observed the pollen

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

tubes make many turns around the papillar cell surface and could not grow toward the style in *L. chinense* pistil, and it might because of the highly expression of plantacyanin in pistil Fig. 3A (Li et al. 2014).

In cluster 4, many genes including *ANXUR1*(*ANX1*) and *NO TRANSMITTING TRACT* (*NTT*) expressed in pollinated pistil and have a peak expression in pollinated pistil after 30 min or 60 min. In Arabidopsis, knockout mutations of the *ANX1* and *ANX2* ceased pollen tube growth in vitro and failed to reach the female gametophytes in vivo. ANX-RLKs constitutively inhibit pollen tube rupture and sperm discharge at the tip of growing pollen tubes to sustain their growth within maternal tissues until they reach the female gametophytes(Boisson-Dernier et al. 2009). In this study, the expression of *ANX1* began to decrease in pistil after 60 min during pollination. Gene *NTT* encoding a C2H2/C2HC zinc finger transcription factor specifically expressed in the transmitting tract. The decreased expression of this gene will result in pollen tubes grow more slowly or terminate prematurely (Crawford et al. 2007). In our previous study, we found that the pollen tube tip cannot reach to the ovules after one hour pollination. The decreased expression of *ANX1* and *NTT* in pistil might cease the pollen tube growth and cause the low seed setting.

3.6 Genes involved in double fertilization process

Except genes involved in the regulation of pollen tube growth processes, some genes involved in double fertilization process were also identified. Two *EMBRYO-DEFECTIVE* (*EMB*) genes (*EMB9* and *EMB2107*) decreased the expression during pollination. A larger number of *EMB* genes with an essential function were required throughout the life cycle in plant. It showed that more than 400 genes required for embryo development in Arabidopsis, and some embryo-defective genes were studied through T-DNA mutants (Devic 2008; Meinke et al. 2009; Muralla et al. 2011). In previous study, *EMB9* and *EMB2107* were classed into preglobular related genes group which were involved in the development of zygotic embryo in early stage (Muralla et al. 2011). Three *UNFERTILIZED EMBRYO*

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

SAC (*UNE*) genes including *UNE10, UNE12, and UNE15* were identified in *L. chinense* pistil during pollination. Down expression of genes *UNE10* and *UNE12* results in unfertilized ovules but normal tube attraction, and the down expression of *UNE15* results in defects in pollen tube attraction in Arabidopsis (Pagnussat et al. 2005). In our study, *UNE10* and *UNE12* were decreased in pistil after 60 min during pollination, and *UNE15* kept increasing during pollination. These results indicated even the pollen tube attraction is well, but there is still problem in double fertilization.

4 Conclusions

Plant sexual reproduction is a complicated biological process with various interactions between pollen and pistil. The interaction between pollen and pistil will determine whether fertilization will be successful or not and thus affect the seed setting. There is only few full seed in aggregate fruit when it matured showed there are some barriers in sexual reproduction process in *L. chinense*. To explore the reason why *L. chinense* has such a low seed setting ratio, transcriptome analysis on pistils of *L. chinense* during pollination were conducted. Finally, a total of 3844 genes which expression fold change during early stage of pollination were higher or lower than 10 were selected as significant differentially expressed genes. Among them, 54 genes of 2440 annotated genes were annotated they might involve in sexual reproduction process including pollen germination, regulation of pollen tube growth, double fertilization forming a zygote and endosperm, and others. The differential expression of genes which involve in the regulation of pollen tube growth process and double fertilization process might be partially causing the low seed setting in *L. chinense*. These results might be helpful to understand why *L. chinense* has such a low seed setting ratio.

Author contribution statement

Conceived and designed the experiments: M.L., K.W., and P.F.Y.; Performed the experiments and analyzed the data: M.L.; Wrote the paper: M.L. and P.F.Y.; Grammar check and suggestion: R.N.D.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

Acknowledgement

We thank Chao Han, Xiaojian Yin, Jiao Deng, Xin Wang, Hui Zhang, Jinlei Fu and Ziyang Fu for assisting in sample collection. We also thank Yuanzhen Liu and Fan Liang in Nextomics Company for data analysis. This project was funded in part by National Natural Science Foundation of China (No. 31300516).

Conflict of interest

The authors affirm that they have no conflicts of interest.

Data archiving statement

The data reported here are deposited and publicly available at the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/). Sequence raw data of *Liriodendron chinense* in fasta format (accession numbers SRR2039541, SRR2148838, SRR2148839, SRR2148840, SRR2148841, SRR2148842) are available at the NCBI.

Supplemental table 1 The comparison of transcriptome data from two research group.

Supplemental table 2 The information of 3844 differentially expressed genes.

Reference

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B Methodological. 57:289-300.

Boisson-Dernier A, Roy S, Kritsas K, Grobei MA, Jaciubek M, Schroeder JI, Grossniklaus U (2009) Disruption of the pollen-expressed FERONIA homologs ANXUR1 and ANXUR2 triggers pollen tube discharge. Development. 136:3279-88.

Cai Z, Penaflor C, Kuehl JV, Leebens-Mack J, Carlson JE, dePamphilis CW, Boore JL, Jansen RK (2006) Complete plastid genome sequences of Drimys, Liriodendron, and Piper: implications for the phylogenetic relationships of magnoliids. BMC evolutionary biology. 6:77.

Chae K, Lord EM (2011) Pollen tube growth and guidance: roles of small, secreted proteins. Annals of Botany. 108:627-636.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

Chen J, Deng F, Li J, Wang QJ (2008) Selective binding of phorbol esters and diacylglycerol by individual C1 domains of the PKD family. The Biochemical journal. 411:333-42.

Chen Z, Liu J, Ng HK, Nadarajah S, Kaufman HL, Yang JY, Deng Y (2011) Statistical methods on detecting differentially expressed genes for RNA-seq data. BMC systems biology. 5 Suppl 3:S1.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 21:3674-6.

Crawford BC, Ditta G, Yanofsky MF (2007) The NTT gene is required for transmitting-tract development in carpels of Arabidopsis thaliana. Current biology : CB. 17:1101-8.

Devic M (2008) The importance of being essential: EMBRYO-DEFECTIVE genes in Arabidopsis. Comptes rendus biologies. 331:726-36.

Fang YM, You LX, Fan RW (1994) Fecundity of natural and cultivated populations of Liriodendron chinense (Hemsl.) Sarg. Journal of Plant Resource and Environment. 3:9-13.

Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic acids research. 36:3420-35.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology. 29:644-652.

Hao RM, He SA, Tang SJ, Wu SP (1995) Geographical distribution of Liriodendron chinense in China and its significance. Journal of Plant Resource and Environment. 4:1-6.

Hoagland DR, Arnon DI (1950) The water-culture method for growing plants without soil. California Agricultural Experiment Station,University of California, Berkeley, CA

Huang H-H, Xu L-L, Tong Z-K, Lin E-P, Liu Q-P, Cheng L-J, Zhu M-Y (2012a) De novo characterization of the Chinese fir (Cunninghamia lanceolata) transcriptome and analysis of candidate genes involved in cellulose and lignin biosynthesis. BMC Genomics. 13:648.

Huang H-H, Xu L-L, Tong Z-K, Lin E-P, Liu Q-P, Cheng L-J, Zhu M-Y (2012b) De novo characterization of the Chinese fir (Cunninghamia lanceolata)

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

transcriptome and analysis of candidate genes involved in cellulose and lignin biosynthesis. BMC Genomics. 13:648.

Huang SQ, Guo YH (2002) Varition of pollination and resource limitation in a low seed-set tree liriodendron chinese. Botanical Journal of the Linnean Society. 140:31-38.

Huang SQ, Guo YH, Pan MQ, Chen JK (1999) Floral Syndrome and Insect Pollination of Liriodendron Chinense. Acta Botanica Sinica. 41:241-248.

Huang SQ, Guo YH, Wu Y, Liu Q, Zhang F, Chen JK (1998) Floral Numerical Variation and Seed Set in Natural Populations of Liriodendron Chinense. Acta Botanica Sinica. 40:22-27.

Iseli C, Jongeneel CV, Bucher P (1999) ESTScan a program for detecting,evaluating, and reconstructing potential coding regions in ESTsequences. Proc Int Conf Intell Syst Mol Biol 138-148.

Korf I, Yandell M, Bedell J (2003) Blast. O'Reilly

Latham RE, Ricklefs RE (1993) continental comparisons of temperate-zone tree species diversity. species diversity in Ecological Communities: Historical and Geographical Perspectives 294-314, University of Chicago Press

Li H, Dong Y, Yang J, Liu X, Wang Y, Yao N, Guan L, Wang N, Wu J, Li X (2012a) De Novo Transcriptome of Safflower and the Identification of Putative Genes for Oleosin and the Biosynthesis of Flavonoids. PloS one. 7(2): e30987

Li HY, Dong YY, Yang J, Liu XM, Wang YF, Yao N, Guan L, Wang N, Wu JY, Li XK (2012b) De Novo Transcriptome of Safflower and the Identification of Putative Genes for Oleosin and the Biosynthesis of Flavonoids. PLOS ONE. 7:e:30987.

Li M, Wang K, Wang X, Yang P (2014) Morphological and proteomic analysis reveal the role of pistil under pollination in Liriodendron chinense (Hemsl.) Sarg. PLOS ONE. 9:e99970.

Li T, Chen J, Qiu S, Zhang Y, Wang P, Yang L, Lu Y, Shi J (2012c) Deep Sequencing and Microarray Hybridization Identify Conserved and Species-Specific MicroRNAs during Somatic Embryogenesis in Hybrid Yellow Poplar. PLOS ONE. 7(8): e43451. doi:10.1371/journal.pone.0043451

Li X, Sun H, Pei J, Dong Y, Wang F, Chen H, Sun Y, Wang N, Li H, Li Y (2012d) De novo sequencing and comparative analysis of the blueberry transcriptome to discover putative genes related to antioxidants. Gene. 511:54-61.

Liang H, Carlson JE, Leebens-Mack JH, Wall PK, Mueller LA, Buzgo M, Landherr LL, Hu Y, DiLoreto DS, Ilut DC, Field D, Tanksley SD, Ma H, dePamphilis CW (2007a) An EST database for Liriodendron tulipifera L. floral buds: the

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

first EST resource for functional and comparative genomics in Liriodendron. Tree Genetics & Genomes. 4:419-433.

Liang HY, Ayyampalayam S, Wickett N, Barakat A, Xu Y, Landherr L, Ralph PE, Jiao YN, Xu T, Schlarbaum SE, Ma H, Leebens-Mack JH, dePamphilis CW (2011) Generation of a large-scale genomic resource for functional and comparative genomics in Liriodendron tulipifera L. Tree Genetics & Genomes. 7:941-954.

Liang HY, Fang EG, Tomkins JP, Luo MZ, Kudrna D, Kim HR, Arumuganathan K, Zhao SY, Leebens-Mack J, Schlarbaum SE, Banks JA, dePamphilis CW, Mandoli DF, Wing RA, Carlson JE (2007b) Development of a BAC library for yellow-poplar (Liriodendron tulipifera) and the identification of genes associated with flower development and lignin biosynthesis. Tree Genetics & Genomes. 3:215-225.

Liu S, Li W, Wu Y, Chen C, Lei J (2013) De novo transcriptome assembly in chili pepper (Capsicum frutescens) to identify genes involved in the biosynthesis of capsaicinoids. PLOS ONE. 8:e48156.

Meinke D, Sweeney C, Muralla R (2009) Integrating the Genetic and Physical Maps of Arabidopsis thaliana: Identification of Mapped Alleles of Cloned Essential (EMB) Genes. PLOS one. 4(10): e7386

Meng AP, Wang HC, Li JQ, SiMa YK (2006) A karyomorphological study of 40 species in 11 genera of the Magnoliaceae from China. Acta Phytotaxonomica Sinica. 44:47.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature methods. 5:621-8.

Muralla R, Lloyd J, Meinke D (2011) Molecular Foundations of Reproductive Lethality in Arabidopsis thaliana. PLOS one. 6(12): e28398.

Nasrallah JB, Kao TH, Chen CH, Goldberg ML, Nasrallah ME (1987) Amino-acid sequence of glycoproteins encoded by three alleles of the S locus of Brassica oleracea. Nature. 326:617-619.

Pagnussat GC, Yu HJ, Ngo QA, Rajani S, Mayalagu S, Johnson CS, Capron A, Xie LF, Ye D, Sundaresan V (2005) Genetic and molecular identification of genes required for female gametophyte development and function in Arabidopsis. Development. 132:603-14.

Parks CR, Wendel JF (1990) Molecular Divergence Between Asian and North American Species of Liriodendron (Magnoliaceae) with Implications for Interpretation of Fossil Floras. Botanical Society of America. 77:1243-1256.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

Qin H, Li BY (1996) Effects of female gametophyte abortion on the reproduction of Liriodendron chinense populations. Journal of Plant Resource and Environment. 5:1-5.

Qin Y, Leydon AR, Manziello A, Pandey R, Mount D, Denic S, Vasic B, Johnson MA, Palanivelu R (2009) Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. PLOS Genetics. 5:e1000621.

Sakamoto K, Kusaba M, Nishio T (1998) Polymorphism of the S-locus glycoprotein gene (SLG) and the S-locus related gene (SLR1) in Raphanus sativus L. and self-incompatible ornamental plants in the Brassicaceae. Molecular Genetics and Genomics. 258:397-403.

Wang K, Li M, Gao F, Li S, Zhu Y, Yang P (2012a) Identification of Conserved and Novel microRNAs from Liriodendron Chinense floral tissues. PLOS ONE. 7(9):e44696. doi:10.1371/journal.pone.0044696

Wang S-Q, Shi D-Q, Long Y-P, Liu J, Yang W-C (2012b) GAMETOPHYTE DEFECTIVE 1, a Putative Subunit of RNases P/MRP, Is Essential for Female Gametogenesis and Male Competence in Arabidopsis. pLOS one. 7(4): e33595. d

Wang S, Xie Y (1992) Red list of endangered plants in China. Science Press, Beijing, China.

Wang Y, Pan Y, Liu Z, Zhu X, Zhai L, Xu L, Yu R, Gong Y, Liu L (2013a) De novo transcriptome sequencing of radish (Raphanus sativus L.) and analysis of major genes involved in glucosinolate metabolism. BMC Genomics. 14:836.

Wang Y, Pan Y, Liu Z, Zhu X, Zhai L, Xu L, Yu R, Gong Y, Liu L (2013b) De novo transcriptome sequencing of radish (Raphanus sativus L.) and analysis of major genes involved in glucosinolate metabolism. BMC Genomics. 14:836.

Ward JA, Ponnala L, Weber CA (2012) Strategies for transcriptome analysis in nonmodel plants. American journal of botany. 99:267-76.

Wen J (1999) Evolution of Eastern Asian and Eastern North American Disjunct Distributions in Flowering Plants. Annual Review of Ecology and Systematics. 30:421-455.

Xu M, Sun Y, Li H (2010) EST-SSRs development and paternity analysis for Liriodendron spp. New Forests. 40:361-382.

Yang AH, Zhang JJ, Tian H, Yao XH (2012) Characterization of 39 novel EST-SSR markers for Liriodendron tulipifera and cross-species amplification in L. chinense (Magnoliaceae). American journal of botany. 99:e460-4.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

Yang AH, Zhang JJ, Yao XH, Huang HW (2011) Chloroplast microsatellite markers in Liriodendron tulipifera (Magnoliaceae) and cross-species amplification in L. chinense. American journal of botany. 98:e123-6.

Yang Y, Xu M, Luo Q, Wang J, Li H (2014) De novo transcriptome analysis of Liriodendron chinense petals and leaves by Illumina sequencing. Gene. 534:155-62.

Yao X, Zhang J, Ye Q, Huang H (2007) Characterization of 14 novel microsatellite loci in the endangered Liriodendron chinense (Magnoliaceae) and cross-species amplification in closely related taxa. Conservation Genetics. 9:483-485.

Ye J, Fang L, Zheng H, Zhang Y, Chen J, Zhang Z, Wang J, Li S, Li R, Bolund L (2006) WEGO: a web tool for plotting GO annotations. Nucleic acids research. 34:W293-W297.

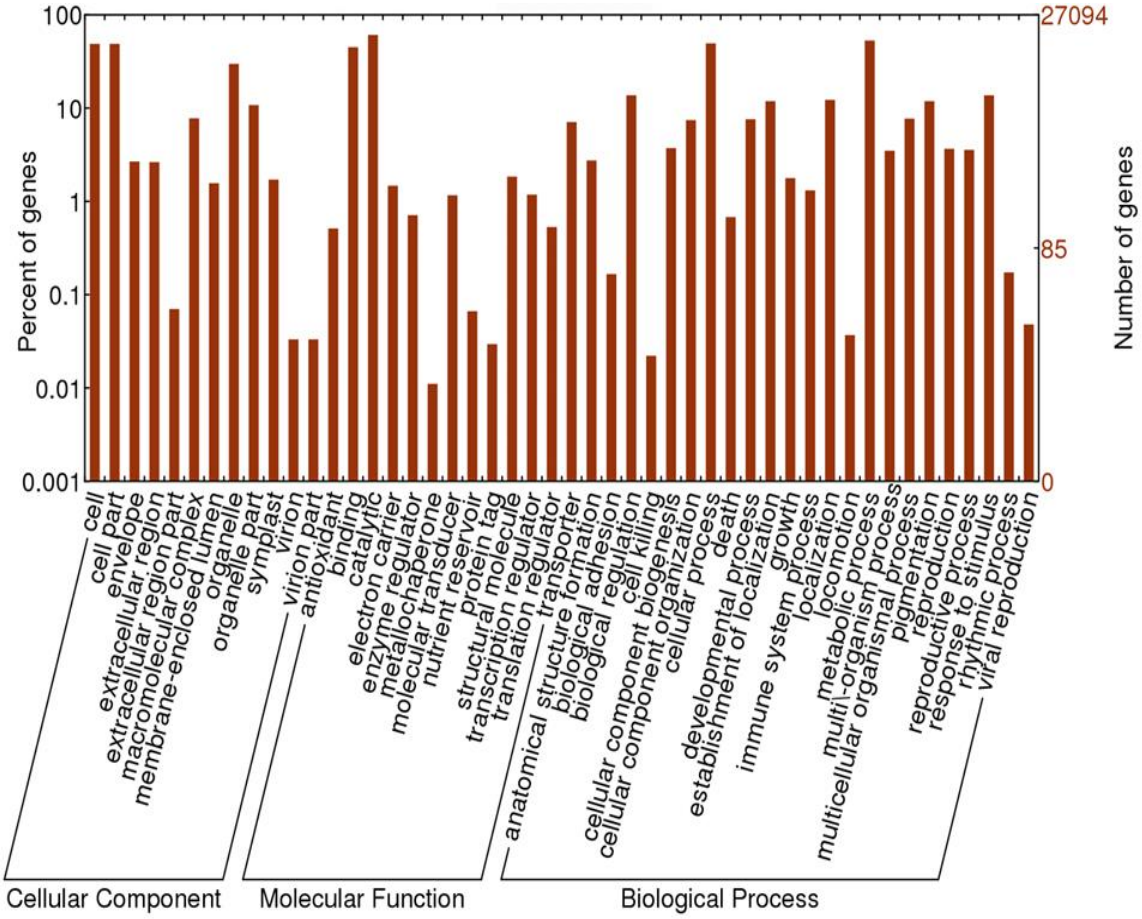
Zhou J, Fan RW (1994) Test of Pollen Quality and Development of Pollen Tubes in Liriodendron. Scientia Silvae Sinicae. 30:405-411.

Zhou J, Fan RW (1999) Pollination of Lriodendron Chinense. Chinese Bulletin of Botany. 16:75-79.

Zhu Q-W, Li H-G (2010) Comparison of genetic diversity among five mating types of Liriodendron revealed by SSR markers. Hereditas (Beijing). 32:183-188.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

604 **Figure legends**



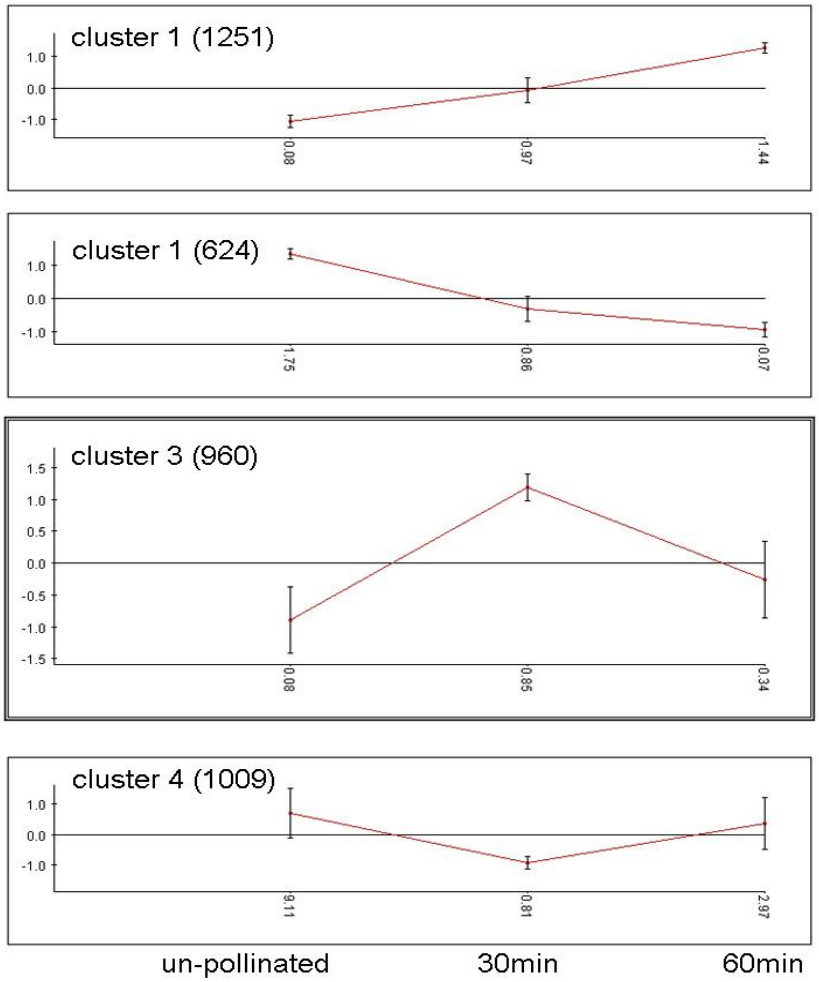
605

606 Figure 1 Gene Ontology classifications of assembled unigenes. The 27094 matched

607 unigenes were classified into 3 functional categories: molecular function, biological

608 process and cellular component.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017



609

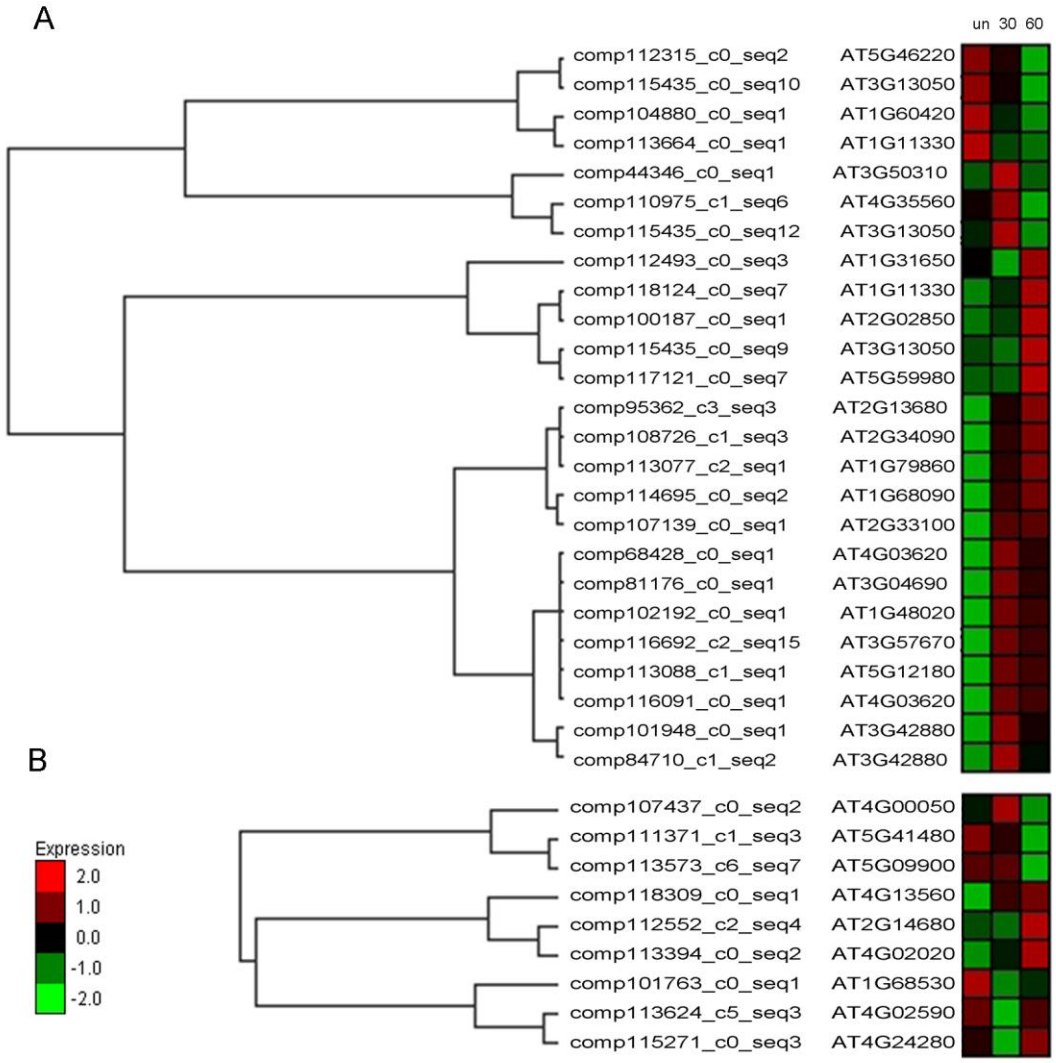
610 Figure 2 Cluster analysis of 3844 differentially expressed genes during early stage of

611 pollination. Un-pollinated, 30min, and 60 min means un-pollinated pistil, pollinated

612 pistil after 30 min, and pollinated pistil after 60 min.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

613



614

615 Figure 3 Cluster analyses of differentially expressed genes which involved in the
616 regulation of pollen tube growth process and double fertilization process during early
617 stage of pollination. Un, 30, and 60 means un-pollinated pistil, pollinated pistil after
618 30 min, and pollinated pistil after 60 min. A means differentially expressed genes
619 which involved in the regulation of pollen tube growth process and B means
620 differentially expressed genes which involved in double fertilization process.
621 Supplementary Figure 1 Comparation of quality of sequencing data between our data
622 and Nanjing's data. a, b were Nanjing's data and c, d were our data. The red and blue
623 lines in a, c means median and average values of quality of sequencing data. The up
624 line and down line of box plots in a, c means maximum and minimum values of

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

quality of sequencing data. The red, blue, green and black lines in b, d means the concentration of T, C, A, and G.

Supplementary Figure 2 Comparation of gene number between our data and Nanjing's data

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

Table 1 Quanlity of sequencing data

Sample	Raw reads	Raw data(M)	Raw Q20(%)	Clean reads	Clean data(M)	Clean Q20(%)	Clean rate (%)	GC(%)
Pistil-0	33149835	6629.97	93.77	32633441	6526.69	93.89	98.44	47
Pistil-30	31284563	6256.91	93.88	30770584	6154.12	94.01	98.36	47
Pistil-60	33499510	6699.9	94.09	32963342	6592.67	94.21	98.4	47
total	97933908	19586.78		96367367	19273.48			47

Table 2 Length distribution of assembled transcripts and unigenes

Nucleotides (bp)	transcript[a]	unigene[b]
200-500bp	151162	
500-1kbp	48106	
1k-2kbp	28406	
>2kbp	15892	
Total	243566	206858
Minimal length(bp)	201	201
Maximal length(bp)	17499	20205
N50 (bp)	1333	963
Average length (bp)	766	646

a means the assembled transcripts in our results

b means assembled unigenes in our results

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

Table 3 Summary of annotated unigenes

	Number of sequence	Percentage (%)
COG	57056	19.8
GO	27094	9.4
KEGG	5399	1.9
NR	92795	32.2
uniport	92363	32.1
ALL	93274	32.4

Table 4 54 differentially expressed genes involved in regulation of pollen tube growth and double fertilization

gene ID	gene accession in arabidopsis	gene name in arabidopsis	function
comp102328_c0_seq1	AT5G09750.1	HEC3	carpel formation
comp113624_c5_seq3	AT4G02590.2	UNE12 UNFERTILIZED EMBRYO SAC 12	double fertilization forming a zygote and endosperm
comp118309_c0_seq1	AT4G13560.1	UNE15 UNFERTILIZED EMBRYO SAC 15	double fertilization forming a zygote and endosperm
comp107437_c0_seq2	AT4G00050.1	UNE10 UNFERTILIZED EMBRYO SAC 10	double fertilization forming a zygote and endosperm
comp115919_c1_seq2	AT1G49540.1	ELP2, AtELP2 | elongator protein 2	embryo sac egg cell differentiation
comp101763_c0_seq1	AT1G68530.1	POLLEN-PISTIL INCOMPATIBILITY 1	embryonic development ending in seed dormancy
comp111371_c1_seq3	AT5G41480.1	EMB9, GLA1,	embryonic development ending in seed dormancy
comp112552_c2_seq4	AT2G14680.3	MEE13 | myosin heavy chain-related	embryonic development ending in seed dormancy
comp113573_c6_seq7	AT5G09900.2	EMB2107, RPN5A, MSA	embryonic development ending in seed dormancy
comp113394_c0_seq2	AT4G02020.1	EZA1, SWN, SDG10	endosperm development
comp102502_c1_seq1	AT2G38110.1	ATGPAT6, GPAT6	flower development

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

comp111300_c1_seq1	AT3G61250.1	AtMYB17, MYB17	flower development
comp112296_c2_seq1	AT1G52150.1	ATHB-15, ATHB15,	flower morphogenesis
comp108285_c2_seq1	AT3G59030.1	TT12, ATTT12 | MATE efflux family protein	maintenance of seed dormancy
comp115483_c0_seq9	AT4G21270.1	ATK1, KATA, KATAP | kinesin 1	multicellular organism reproduction
comp115271_c0_seq3	AT4G24280.1	cpHsc70-1	ovule development
comp117463_c2_seq4	AT2G44680.1	CKB4 | casein kinase II beta subunit 4	photoperiodism, flowering
comp95362_c3_seq3	AT2G13680.1	CALS5, GLS2, ATGSL02 | callose synthase 5	pollen germination
comp107139_c0_seq1	AT2G33100.1	ATCSLD1, CSLD1 | cellulose synthase-like	pollen germination
comp44346_c0_seq1	AT3G50310.1	MAPKKK20	pollen sperm cell differentiation
comp110975_c1_seq6	AT4G35560.1	Transducin/WD40 repeat-like superfamily	pollen sperm cell differentiation
comp108726_c1_seq3	AT2G34090.1	MEE18 | maternal effect embryo arrest 18	pollen tube development
comp117121_c0_seq7	AT5G59980.2	GAMETOPHYTE DEFECTIVE 1	pollen tube development
comp81176_c0_seq1	AT3G04690.1	ANX1 | Malectin/receptor-like protein kinase	pollen tube growth
comp102192_c0_seq1	AT1G48020.1	PMEI1, ATPMEI1	pollen tube growth
comp113077_c2_seq1	AT1G79860.1	ATROPGEF12, ROPGEF12, MEE64	pollen tube growth
comp116091_c0_seq1	AT4G03620.1	myosin heavy chain-related	pollen tube growth
comp116692_c2_seq15	AT3G57670.1	NTT, WIP2 | C2H2-type zinc finger family	pollen tube growth
comp68428_c0_seq1	AT4G03620.1	myosin heavy chain-related	pollen tube growth
comp112315_c0_seq2	AT5G46220.1	Protein of unknown function (DUF616)	pollen tube growth
comp114695_c0_seq2	AT1G68090.1	ANNAT5, ANN5 | annexin 5	pollen tube growth
comp115435_c0_seq9	AT3G13050.1	Major facilitator superfamily protein	pollen tube growth
comp115435_c0_seq10	AT3G13050.1	Major facilitator superfamily protein	pollen tube growth
comp115435_c0_seq12	AT3G13050.1	Major facilitator superfamily protein	pollen tube growth
comp104880_c0_seq1	AT1G60420.1	DC1 domain-containing protein	pollen tube guidance
comp100187_c0_seq1	AT2G02850.1	ARPN | plantacyanin	pollination

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017

comp113664_c0_seq1	AT1G11330.1	S-locus lectin protein kinase family protein	recognition of pollen
comp118124_c0_seq7	AT1G11330.1	S-locus lectin protein kinase family protein	recognition of pollen
comp84710_c1_seq2	AT3G42880.1	Leucine-rich repeat protein kinase family	regulation of pollen tube growth
comp101948_c0_seq1	AT3G42880.1	Leucine-rich repeat protein kinase family	regulation of pollen tube growth
comp112493_c0_seq3	AT1G31650.1	ATROPGEF14, ROPGEF14	regulation of pollen tube growth
comp113088_c1_seq1	AT5G12180.1	CPK17 | calcium-dependent protein kinase	regulation of pollen tube growth
comp113744_c3_seq13	AT5G38030.1	MATE efflux family protein	ripening
comp94208_c0_seq1	AT2G41850.1	PGAZAT, ADPG2	ripening
comp113488_c2_seq2	AT5G17420.1	IRX3, CESA7, ATCESA7, MUR10	seed coat development
comp107765_c0_seq2	AT2G18390.1	TTN5, HAL, ARL2, ATARLC1	seed development
comp114733_c1_seq9	AT1G28440.1	HSL1 | HAESA-like 1	seed dormancy
comp114470_c0_seq8	AT5G51760.1	AHG1 | Protein phosphatase 2C family protein	seed dormancy
comp109912_c0_seq1	AT4G17030.1	ATEXLB1, EXPR, AT-EXPR, ATEXPR1	sexual reproduction
comp118159_c0_seq1	AT1G79280.1	NUA, AtTPR | nuclear pore anchor	stamen development
comp118159_c0_seq3	AT1G79280.1	NUA, AtTPR | nuclear pore anchor	stamen development
comp113694_c0_seq1	AT3G44260.1	Polynucleotidyl transferase	vegetative to reproductive phase transition
comp115565_c0_seq1	AT3G44260.1	Polynucleotidyl transferase	vegetative to reproductive phase transition
comp116715_c1_seq1	AT3G50530.1	CRK | CDPK-related kinase	vegetative to reproductive phase transition

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.2803v1 | CC BY 4.0 Open Access | rec: 15 Feb 2017, publ: 15 Feb 2017