



Article

Pinus massoniana Introgression Hybrids Display Differential Expression of Reproductive Genes

Jiaxing Mo ^{1,2} , Jin Xu ^{1,2,*} , Yuting Cao ^{1,2}, Liwei Yang ^{1,2}, Tongming Yin ^{1,2}, Hui Hua ^{1,2}, Hui Zhao ^{1,2}, Zhenhao Guo ^{1,2}, Junjie Yang ^{1,2} and Jisen Shi ^{1,2}

¹ Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, China; jx_mo@njfu.edu.cn (J.M.); 18205190992@163.com (Y.C.); yanglw@njfu.edu.cn (L.Y.); tmyin@njfu.edu.cn (T.Y.); 13218098608@163.com (H.H.); luoyusheng6@163.com (H.Z.); 18915019168@163.com (Z.G.); 18260077751@163.com (J.Y.); jshi@njfu.edu.cn (J.S.)

² College of Forestry, Nanjing Forestry University, Nanjing 210037, China

* Correspondence: xjinhsh@njfu.edu.cn; Tel.: +86-25-8542-7319

Received: 23 January 2019; Accepted: 28 February 2019; Published: 5 March 2019



Abstract: *Pinus massoniana* and *P. hwangshanensis* are two conifer species located in southern China, which are of both economic and ornamental value. Around the middle and lower reaches of the Yangtze River, *P. massoniana* occurs mainly at altitudes below 700 m, while *P. hwangshanensis* can be found above 900 m. At altitudes where the distribution of both pines overlaps, a natural introgression hybrid exists, which we will further refer to as the Z pine. This pine has a morphological character that shares attributes of both *P. massoniana* and *P. hwangshanensis*. However, compared to the other two pines, its reproductive structure, the pinecone, has an ultra-low ripening rate with seeds that germinate poorly. In this study, we aimed to find the reason for the impaired cone maturation by comparing transcriptome libraries of *P. massoniana* and Z pine cones at seven successive growth stages. After sequencing and assembly, we obtained unigenes and then annotated them against NCBI's non-redundant nucleotide and protein sequences, Swiss-Prot, Clusters of Orthologous Groups, Gene Ontology and KEGG Orthology databases. Gene expression levels were estimated and differentially expressed genes (DEGs) of the two pines were mined and analyzed. We found that several of them indeed relate to reproductive process. At every growth stage, these genes are expressed at a higher level in *P. massoniana* than in the Z pine. These data provide insight into understanding which molecular mechanisms are altered between *P. massoniana* and the Z pine that might cause changes in the reproductive process.

Keywords: *Pinus massoniana*; introgression hybrid; RNA sequencing; DEGs; reproduction

1. Introduction

Gymnosperms have their own unique way of reproduction. Microspores grow into pollen carrying sperm cells, while megaspores develop into the megagametophyte. The archegonium forms during development of the megagametophyte. Then, pollen enter ovule through micropyle, they move towards the egg by way of extending a pollen tube, after which fertilization occurs. Finally, the embryo is formed and develops into a gymnosperm seed.

Conifers possess a series of properties that makes exploring their molecular biology through a genomics approach challenging, such as a long life cycle, a reproductive process lasting months or even years, a gigantic genome size and so forth [1,2]. We aimed to explore the molecular mechanism of conifer reproduction by generating transcriptome data through RNA-seq, of successive stages of

the developing pinecone. RNA-seq technology has advanced to the stage where it is highly efficient, sensitive and accurate.

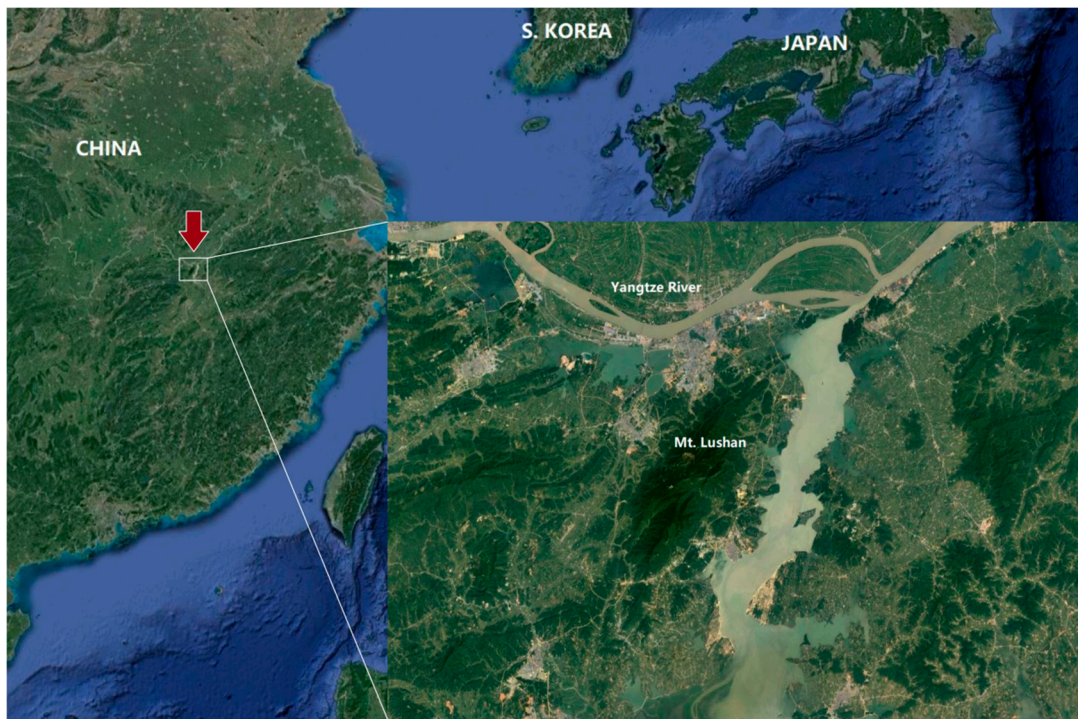
By studying the expression dynamics of differentially expressed genes (DEGs) in pinecones, we sought to gain insight into the relevant genes that control the reproductive process of conifers. Previous empirical studies have suggested several genes linked to reproduction, such as *DAL* [3], *MADS-box* [4,5], *MYB* [6] and *MSI* [7] and so forth. However, studies of the determinants of development and regulation of reproduction have concentrated on model angiosperm species so far, while gymnosperms remain largely understudied.

We centered our studies around two types of conifers, naturally occurring in China. First, *Pinus massoniana*, also known as the masson pine, is an economically important conifer species. It is mainly distributed in various southern Chinese provinces, at altitudes below 700 m around the middle and lower reaches of the Yangtze River. It offers wood and pulp for manufacturing furniture and paper and also supplies natural resin, which can be further manufactured into 'resin,' a crucial product used in the maintenance of instrument strings and as an ingredient in medicine. Furthermore, *P. massoniana* fulfills a significant ecological role by replacing or compensating natural forest destruction due to its fast growth and abundant biomass.

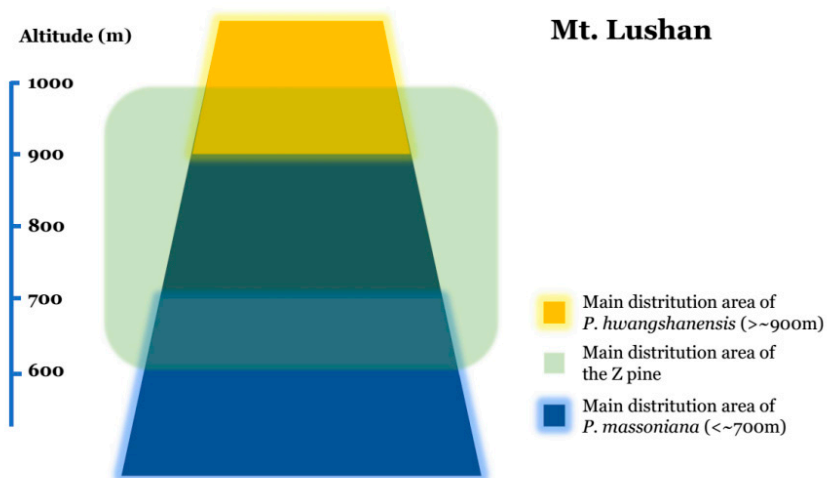
The second, *Pinus hwangshanensis*, primarily grows in southeastern China. It grows most abundantly at an altitude above 900 m around the middle and lower reaches of the Yangtze River. Growing at a higher altitude limits its speed of growth, as well as accumulation of biomass and the resin compared to *P. massoniana*. *P. hwangshanensis* is often viewed as luxurious and graceful, making it an ideal ornamental tree.

Mountain Lushan (Figure 1a) is located within the distribution area of both conifer species. Its peak has an altitude of 1474 m and it supports both *P. massoniana* and *P. hwangshanensis* vegetation at the afore mentioned altitudes. Where *P. massoniana* and *P. hwangshanensis* distribution overlaps, a natural introgression hybrid of both species occurs, sharing phenotypic characters of both its parents (Figure 1b) [8,9]. Due to the hybrid not being named yet, we refer to it as the 'Z pine' in this article. The Z pine has an extremely low germination and ripening rate compared to both *P. massoniana* and *P. hwangshanensis* [10]. These characters could indicate that the Z pine displays genetic incompatibilities during fertilization and/or even embryonic development. What causes this phenomenon is still unknown.

In this study, we collected seven successive development stages of open-pollinated cones *P. massoniana* and the Z pine, respectively. Then we characterized the transcriptome of these cones using Illumina high-throughput sequencing technology and forty two cDNA libraries were constructed. A series of experiments was performed to mine candidate genes, focusing on differential expression patterns between these two species. Moreover, differentially expressed genes related to fertilization and embryonic development were determined and analyzed in both taxa. This study could help explaining the defect of the Z pine of its unusual low ripening rate and germination rate, comparing to *P. massoniana* and may provide an approach to understanding difference between species and its introgressive hybrid at the transcriptome level.



(a)



(b)

Figure 1. (a) Location of Mountain Lushan. Mt. Lushan lies in Jiujiang City, Jiangxi Province, China. The Great Han Yang Peak, the highest point of the mountain, is 1474 m high. The local climate is humid, subtropical; (b) Schematic diagram of main distribution area of *P. hwangshanensis*, the Z pine and *P. massoniana* on Mt. Lushan.

2. Materials and Methods

2.1. Sample Collection

Differently staged, openly pollinated cones of *P. massoniana* and the Z pine were collected on Mt. Lushan, Jiujiang, China (Table 1, Figure 2). Due to complex environment of forest land and wind-pollinated way of pine, sampled individuals (especially the Z pine) may possess different level of

introgressive background. In order to minimize the different between them, we assigned five maternal trees of each taxa (*P. massoniana* or the Z pine) in its sample plot. Experience tells us that *P. massoniana* carries out its pollination at around 10 April while the Z pine at around 20 April at sample plots. Aware of this condition, we conducted our first sample collection on 27 April (Table 1), when both *P. massoniana* and the Z pine are already pollinated. The cones were packed with aluminum-foil shortly after collection and then immediately submerged in liquid nitrogen, after which they were stored in a $-80\text{ }^{\circ}\text{C}$ freezer until RNA extraction.

Table 1. Information on the geographical sites of sample collection of *P. massoniana* and the Z pine.

Species	Sample Code							Location		
	Open Pollinated Cones of the 1st Year				Open Pollinated Cones of the 2nd Year			Longitude (E)	Latitude (N)	Altitude (m)
	APR. 27	MAY 25	JUL. 27	SEPT. 25	APR. 27	JUL. 27	SEPT. 25			
<i>P. massoniana</i>	MA	MB	MC	MD	ME	MF	MG	116.04	29.50	78
the Z pine	ZA	ZB	ZC	ZD	ZE	ZF	ZG	115.98	29.54	730

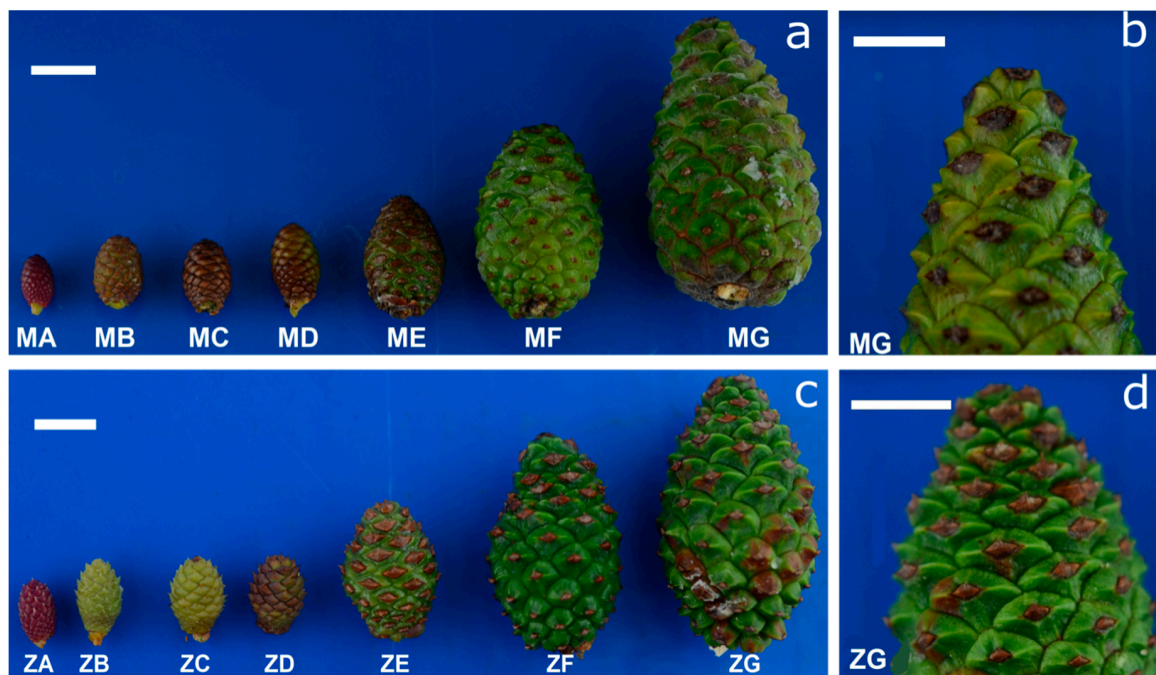


Figure 2. Successively staged pine cones of *P. massoniana* (a) and the Z pine (c). Details of mature cones of MG (b) and ZG (d) stages are shown. The scale shields of *P. massoniana* are flat or slightly bulged, the transverse ridge is not very obvious, and the scale umbilical has no thorn. The scale shield of the Z pine is bulged, the transverse ridge is obvious, and the scale umbilical is thorny. Codenames for collected cones are explained in Table 1. (Scale = 10 mm.).

2.2. RNA Extraction and Sequencing

We randomly collected three to six cones for RNA isolation of each sample code to make sure that these cones and their RNA could be representative. Cones were taken from the $-80\text{ }^{\circ}\text{C}$ freezer and briefly re-frozen in liquid nitrogen to further weaken tissue, cut into pieces, after which sections containing ovules (or seeds) were collected and crushed. RNA was extracted from each sample using the Biotek Plant RNA Extraction Kit (Beijing, China). Three replications of RNA were extracted for each sample code. Purity and quality of the RNA samples was checked respectively by measuring 260 nm/280 nm UV absorption values with a Nanodrop 2000 (Thermo Fisher Scientific, Waltham, MA,

USA) and examining the RIN (RNA Integrity Number) with an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

After RNA extraction, magnetic oligo (dT) beads were used to purify mRNA, which was then collected using RNeasy RNA reagent. The mRNA was then cut into small fragments using the RNA Fragment Reagent (Illumina, San Diego, CA, USA) and subsequently cleaned using an RNeasy RNA Cleaning Kit (Qiagen, Germany). First-strand cDNA was then synthesized using MMLV reverse transcriptase (Takara, Japan), while second-strand cDNA synthesis was performed using DNA Polymerase I and RNase H. cDNA was finally sequenced on an Illumina HiSeq X Ten (Illumina, USA). The sequencing raw data were submitted to the NCBI Short Reads Archive (SRA) database under the BioProject accession number PRJNA482692.

2.3. Data Processing and Assembly

The raw RNA-seq data was screened by removing adaptor and low quality sequences using Trimmomatic [11]. Qualified reads were assembled into non-redundant transcripts by Trinity [12,13], using the following parameters: `-min_contig_length 200 -min_kmer_cov 2 -min_glue 3 -seqType fq`, other parameters use default settings. Contigs were assembled by Trinity, gathered and processed by TGICL [14], the parameters of TGICL are as follows: `-l 40 -c 10 -v 25 -O -repeat_stringency 0.95 -minmatch 35 -minscore 35`. After which unigenes were collected.

2.4. Functional Annotation

Unigenes were generally annotated by aligning against SwissProt (Release-2016_07) [15], NCBI non-redundant protein sequences (Nr, Release-20160314), NCBI non-redundant nucleotide sequences (Nt, Release-20140514) [16], Kyoto Encyclopedia of Genes and Genomes (KEGG, Release 59.3) [17,18], Cluster of Orthologous Groups of proteins (COG, Release-20090331) [19] and Gene Ontology (GO) [20] databases, selecting the most likely annotations. Blast2GO (v2.5.0) [21] was conducted as GO annotation tool against GO database (Release-201604) under default settings. Furthermore, analysis of GO function and KEGG pathways of differentially expressed genes was performed.

2.5. Differentially Expressed Genes (DEGs) and Gene Expression Pattern Analysis

Calculation of unigene read counts was performed using RNA-Seq by Expectation-Maximization (RSEM) software [22]. RSEM results were transformed into FPKM [23] values (expected number of Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced), commonly used for measuring gene expression levels. DESeq was used to determine differentially expressed genes of different transcript libraries [24]. Differentially expressed genes were assigned based on a threshold value of FDR (false discovery rate) ≤ 0.001 and $|\log_2\text{Ratio}| \geq 2$. Gene expression patterns of *P. massoniana* and the Z pine were assembled by Short Time-series Expression Miner (STEM) [25].

2.6. Validation by Quantitative Real-Time PCR (qRT-PCR)

Quantitative real-time PCR was applied for validating differentially expressed genes detected by our RNA-seq analysis. Primers were designed using the NCBI Primer-Blast Tool [26] and synthesized by Generay Biotech Co., Ltd. (Shanghai, China). cDNA samples for qRT-PCR were synthesized using the Vazyme HiScript II Q RT SuperMix for qPCR (Nanjing, China). qRT-PCR was carried out using an Applied Biosystems 7500 PCR cycler (Thermo Fisher Scientific Corporation, CA, USA) and Vazyme ChamQ SYBR qPCR Master Mix (Nanjing, China) as reaction reagent kit. Each sample was run in triplicate, with samples having a final volume of 20 μL : containing 10 μL of ChamQ SYBR qPCR Master Mix (2 \times), 0.4 μL of each primer, 2 μL of cDNA and 7.2 μL of ddH₂O. The reaction program was according to standard product instructions. An *Actin* gene that was discovered from RNA-seq data (Unigene69821_All) was utilized as reference gene. The qRT-PCR data was analyzed with the $2^{-\Delta\Delta\text{Ct}}$ method [27].

3. Results

3.1. Illumina Sequencing and Assembly

A total of 160 Gb raw data was obtained. It has an average of 53 million reads per library. We evaluated the quality of the original and clean sequencing data of all samples. The Q30 value has a range of 88.49~92.33 % for the original sequencing data (Table S1) and a range of 91.75~94.31% for the clean data (Table S2), indicating that this data set is ready for further assembly. All transcripts that we obtained from the staged pine cones of our two conifer species were assembled into 93,291 unigenes (see Materials and methods for details, Table 2). About 39.88% of them exceeded 2 kb in length, while 37.02% unigenes have a length from 1 kb to 2 kb and 23.1% were 100 bp to 1 kb (Figure S1). The average length of unigenes is 1987 nt, while N50 is 2494 nt.

Table 2. Number and length of unigenes.

Sample	Total Number	Total Length (nt)	Mean Length (nt)	N50
MA	77,051	89,621,989	1163	1875
MB	75,111	86,567,730	1153	1838
MC	85,801	96,687,090	1127	1833
MD	78,553	88,590,120	1128	1834
ME	78,223	87,067,948	1113	1813
MF	76,159	84,583,542	1111	1800
MG	100,035	87,128,148	871	1400
ZA	68,015	80,547,061	1184	1893
ZB	76,206	88,922,013	1167	1882
ZC	81,570	90,787,807	1113	1824
ZD	74,227	84,044,527	1132	1823
ZE	66,146	76,428,873	1155	1833
ZF	73,004	79,225,689	1085	1789
ZG	85,392	86,084,303	1008	1708
All	93,291	185,359,493	1987 ¹	2494 ²

^{1,2} When all samples are assembled, they would express much higher abundance than single samples, therefore the data of 'All' is usually higher than others.

3.2. Functional Annotation of *P. massoniana* and Z Pine Unigenes

The assembled unigenes were annotated against the SwissProt, NCBI non-redundant protein and nucleotide sequences (Nr and Nt), Kyoto Encyclopedia of Genes and Genomes (KEGG), Cluster of Orthologous Groups of proteins (COG) and Gene Ontology (GO) databases. A total of 86,006 unigenes were annotated in *P. massoniana* and the Z pine, of which 25,150 unigenes were annotated against all six databases (Figure S2).

We were able to annotate most genes using the Nr database, using sequences from *Picea sitchensis* for the bulk of the annotation (30,943), after which *Amborella trichopoda* (7055) and Indian lotus (4504) provided most annotations, suggesting that these sequenced species are most closely related to *P. massoniana* and the Z pine (Figure S3).

We then used data from the COG, GO and KEGG databases for unigene functional prediction. Using the COG database, 30,017 unigenes could be annotated and were classified into 24 functional categories. The 'general function prediction only' was the most abundant, followed by 'transcription' and 'replication, recombination and repair,' 'function unknown' and 'signal transduction mechanisms' (Figure 3).

A total of 229,950 redundant unigenes (with 38,619 nonredundant unigenes) were annotated into 56 sub-categories under three primary GO categories: biological process, cellular component and molecular function (Figure 4). The top three sub-categories were metabolic process (23,854 unigenes), cellular process (22,439 unigenes) and catalytic activity (20,324 unigenes).

The KEGG classification placed 41,931 unigenes into 5 pathway functional categories (Figure 5): organismal systems (5474 unigenes), metabolism (37,045 unigenes), genetic information processing (9684 unigenes), environmental information processing (2871 unigenes) and cellular processes (1487 unigenes). The top three sub-categories out of a total of 18 were ‘global map,’ ‘environmental adaptation’ and ‘carbohydrate metabolism,’ which contains 14,748, 5350 and 4965 unigenes, respectively.

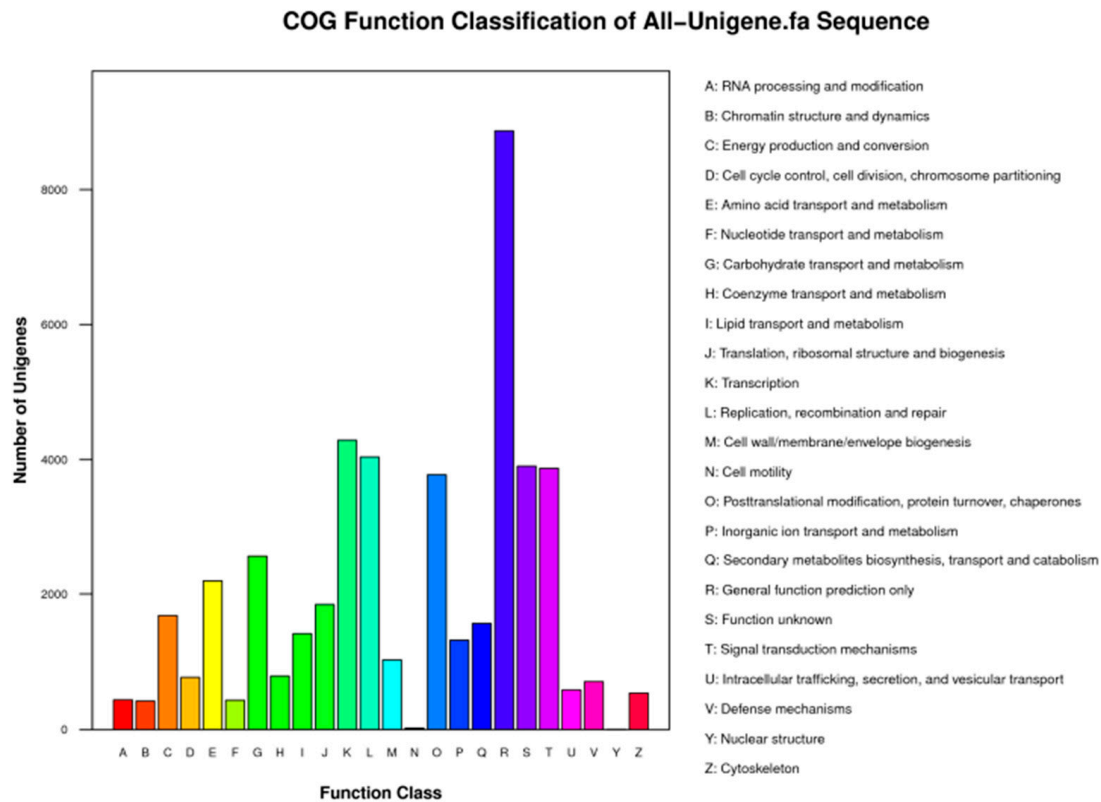


Figure 3. COG functional classification of *P. massoniana* and the Z pine unigenes.

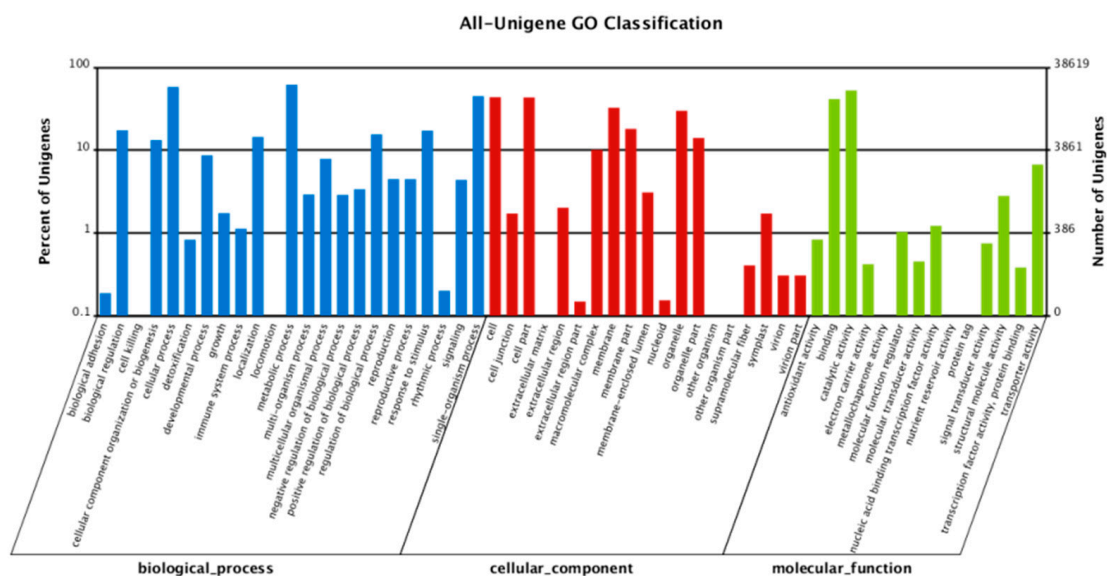


Figure 4. Gene ontology (GO) functional classification of *P. massoniana* and the Z pine unigenes.

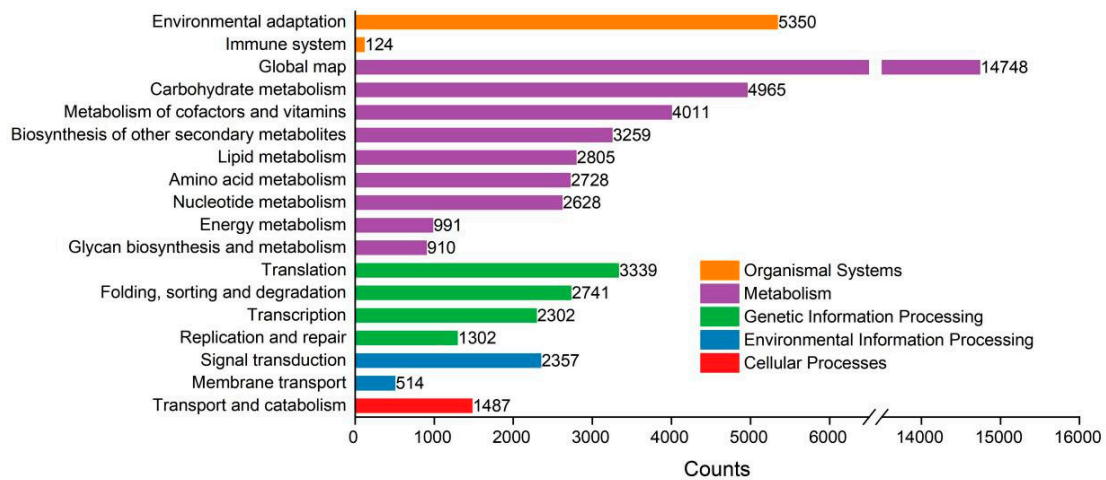


Figure 5. Histogram of the KEGG Pathway classification of *P. massoniana* and the Z pine unigenes.

3.3. Analysis of Expected Number of Fragments per Kilobase of Transcript Sequence per Million Base Pairs Sequenced (FPKM)

FPKM values were calculated using RSEM software. The general density distribution of expression quantity (Figure 6) was analyzed and showed that the average total of expressed mRNAs across all unigenes of *P. massoniana* and the Z pine varies between species and stages.

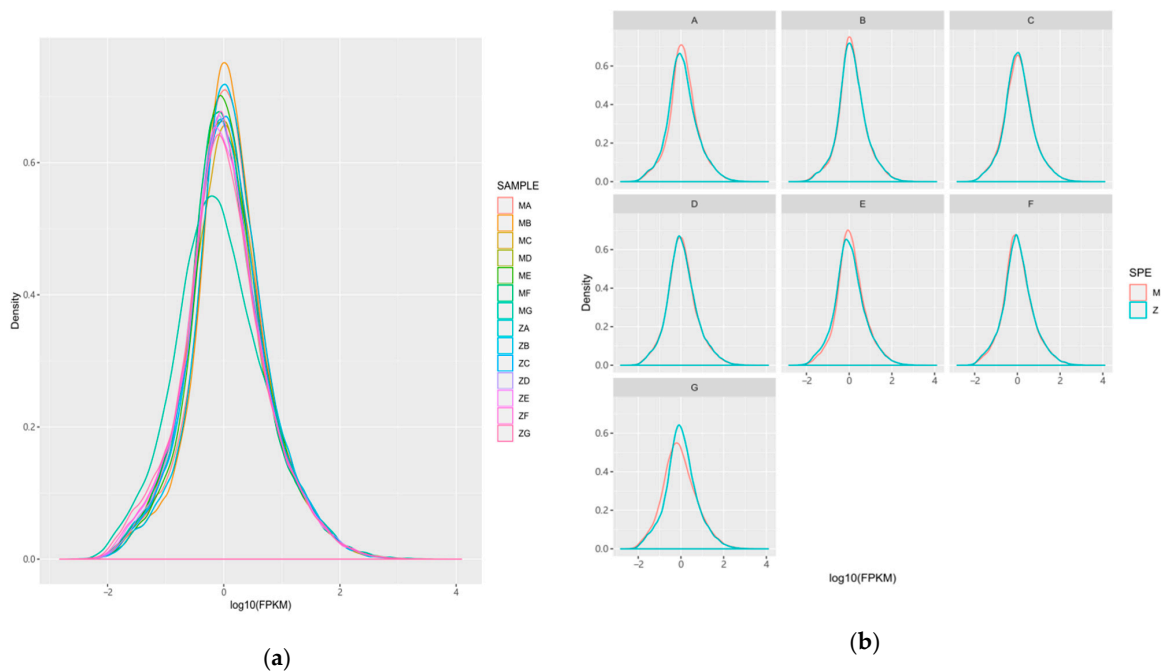


Figure 6. (a) General density distribution of *P. massoniana* and Z pine unigenes by FPKM analysis; (b) Pairwise comparison of general density distribution of same stage *P. massoniana* and Z pine unigenes by FPKM analysis, M: *P. massoniana* (red curve), Z: Z pine (green curve), A~G: the seven successive stages of pinecone development. X-axis: logarithm to base 10 of FPKM, y-axis: density of distribution.

3.4. GO Classification and KEGG Enrichment Assessment of Differentially Expressed Genes (DEGs) at Successive Pinecone Stages

We determined and compared the number of up- and down-regulated genes between the two pine species at the seven different developmental stages of the pine cones collected (Figure 7 and Figure S4).

In comparison group A, C and G, more down-regulated genes were found than up-regulated ones, while in group B, D, E and F, there was more up-regulated genes than down-regulated ones (Figure 7).

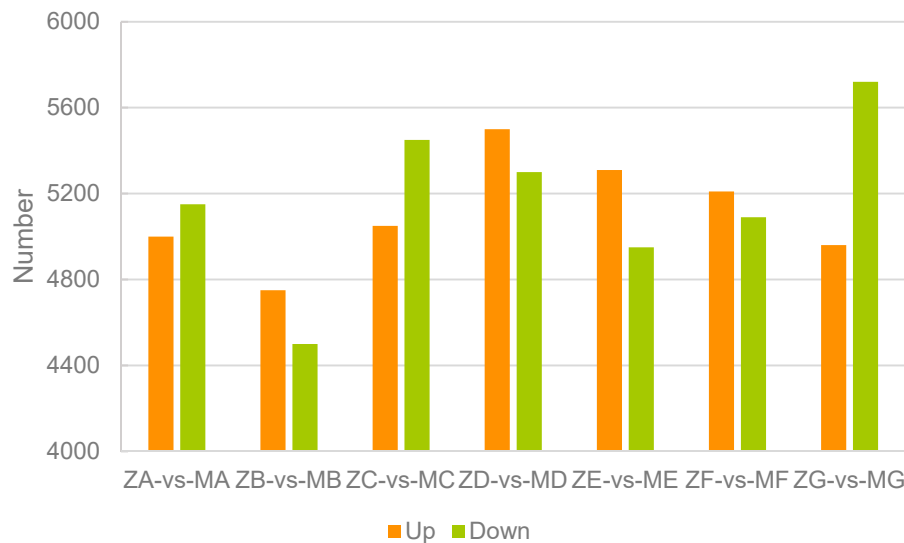


Figure 7. Statistics of differentially expressed genes of same stages of two pines ($FDR \leq 0.001$ and $|\log_2 \text{Ratio}| \geq 2$). Red indicates up-regulated genes while green indicates down-regulated genes. The first two bars show that compare to ZA, MA has more down-regulated genes than up-regulated ones, the rest can be read in this manner.

To gain more insight in differential regulation of genes related to the pinecone reproductive process, we performed GO classification of all DEGs at every developmental stage (Figure S5). The number of differentially expressed unigenes classified in the ‘reproduction’ category under ‘biological process’ at each stage is listed (Figure S6). Three groups displayed more up- than down-regulated genes: ZA-MA, ZD-MD and ZF-MF. While the other four groups showed more down- than up-regulated genes: ZB-MB, ZC-MC, ZE-ME and ZG-MG. The ZG-MG group contains 101 down-regulated genes, almost double the number of up-regulated genes (51 genes).

We then determined whether specific cellular processes are differentially affected at each pinecone stage by performing a KEGG pathway enrichment analysis (Figure S7).

3.5. Quantitative Real-Time PCR Validation

We randomly selected five unigenes for validation of the accuracy of our RNA-seq data set using qRT-PCR. The following unigenes were randomly chosen: Unigene12135_All, Unigene31229_All, Unigene5965_All, Unigene69986_All and Unigene71003_All. We tested cDNA derived from four samples: MA, ME, ZA, ZE, which were collected in different years from the two different species. Details of sequences and primers were list on Tables S3 and S4, respectively. Validation results shows a reliable correlation between RNA-seq and qRT-PCR (Figure S8).

3.6. Temporal Gene Expression Profiles of *P. massoniana* and the Z Pine

We analyzed gene expression dynamics of all unigenes across pinecone developmental stages for both species and clustered these into 49 unique expression profiles. The eighteen most frequently occurring profiles for each species are shown in Figure 8a. Within *P. massoniana*, the top five expression profile types are 16, 10, 31, 34 and 40, with respectively 8416, 7967, 5532, 3291 and 3092 genes showing these expression dynamics. In the Z pine the most frequently occurring profiles are profile types 40, 10, 27, 29 and 44, represented by 5610, 4336, 3441, 3371 and 3361 genes. *P. massoniana* and the Z pine shared fourteen profile types among their respective top 18 profiles (Figure 8a). In addition, we carried

out a profile comparison between *P. massoniana* and the Z pine. Every single profile of *P. massoniana* is listed and similar ones of Z pine are placed on the right of it in Figure 8b.

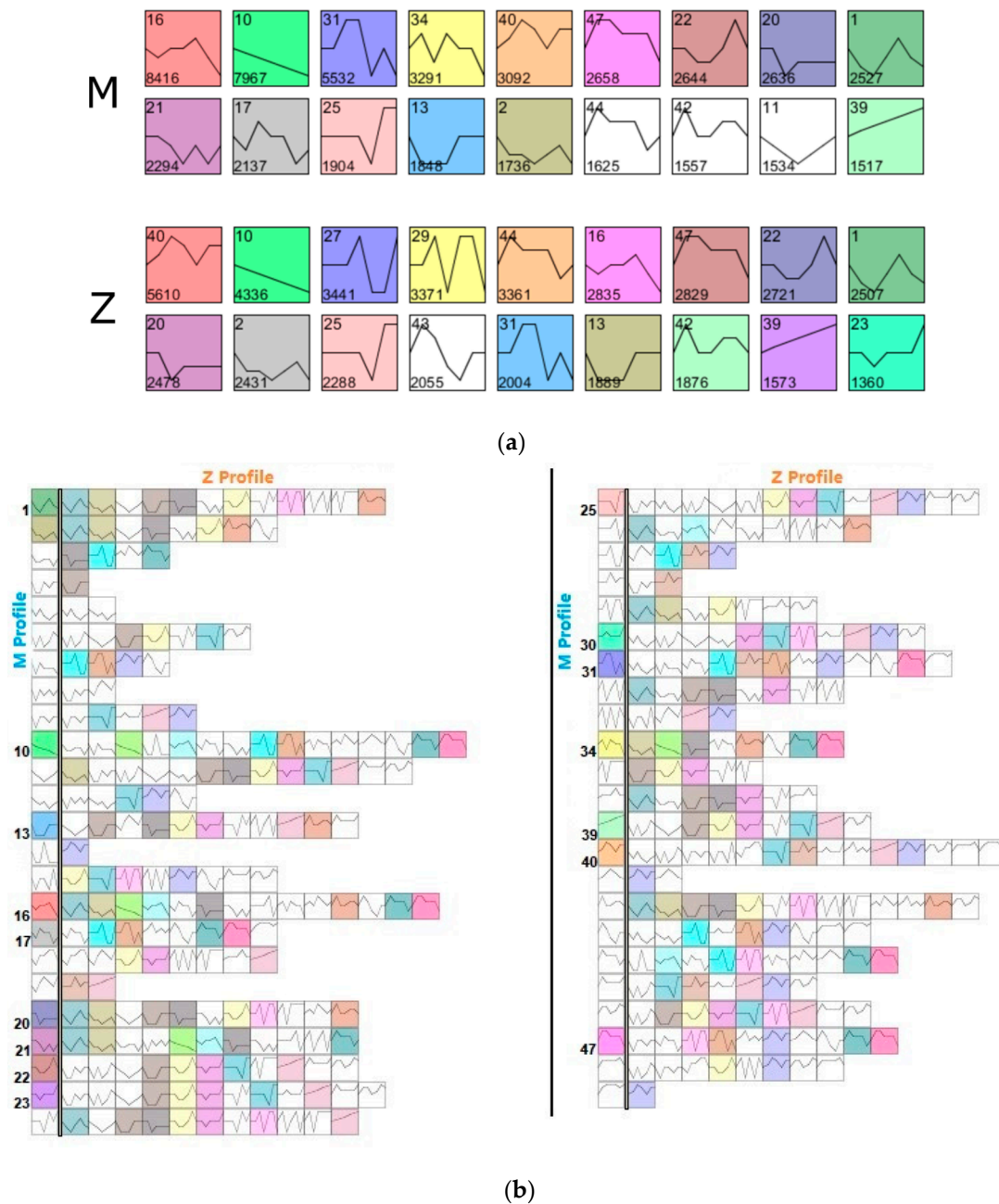


Figure 8. (a) Eighteen expression patterns of *P. massoniana* (M) and the Z pine (Z) based on the highest number of genes having these particular patterns. The number on the upper left of each square indicates profile type, the number on the lower left indicates the number of genes within each profile, the fold line shows the expression pattern, colored squares are significant profiles while white ones are insignificant profiles; (b) All *P. massoniana* (M) patterns compare to its similar ones in the Z pine (Z). The most left column of each part is profiles of *P. massoniana* while their similar Z pine counterparts are listed on the right, significant profiles of M are marked profile numbers and colors.

3.7. Reproductive Genes Are Differentially Expressed between *P. massoniana* and the Z Pine

Next, we aimed to see whether genes related to reproduction might be differentially expressed between *P. massoniana* and the Z pine, potentially explaining the reproductive problems that the Z

pine experiences. We looked for several DEGs involved in processes such as pollen development, pollen exine formation, pollen tube growth and development of the female gametophyte, endosperm, embryo and/or embryo sac according to recent reports (Table S5). Some of these genes show consistently higher expression in *P. massoniana* than in the Z pine, including: *ACA7* (Ca²⁺-ATPase), *MPK4* (mitogen-activated protein kinase), *QRT2* (polygalacturonase), *TKPR1* (tetraketide alpha-pyrone reductase 1), *PI5K* (phosphatidylinositol 4-phosphate 5-kinase), *PMEs* (pectin methylesterase), *SEC6* (exocyst complex component 6), *SEC15* (exocyst complex component 15), *SWK2* (slow walker 2), *PPR* (pentatricopeptide repeat-containing protein), *EMB* (embryo), *LEA* (late embryogenesis abundant protein), *SERK* (somatic embryogenesis receptor kinase), *BLH* (*BEL1-like* homeodomain protein) (Figure 9). Some genes relate to reproduction show a similar expression level in both pines, such as: *SHT* (spermidine hydroxycinnamoyl transferase), *SEC5* (exocyst complex component 5), *SEC8* (exocyst complex component 8), *EYE* (embryo yellow), *EDD1* (embryo defective development 1) and *EDA* (embryo sac development arrest).

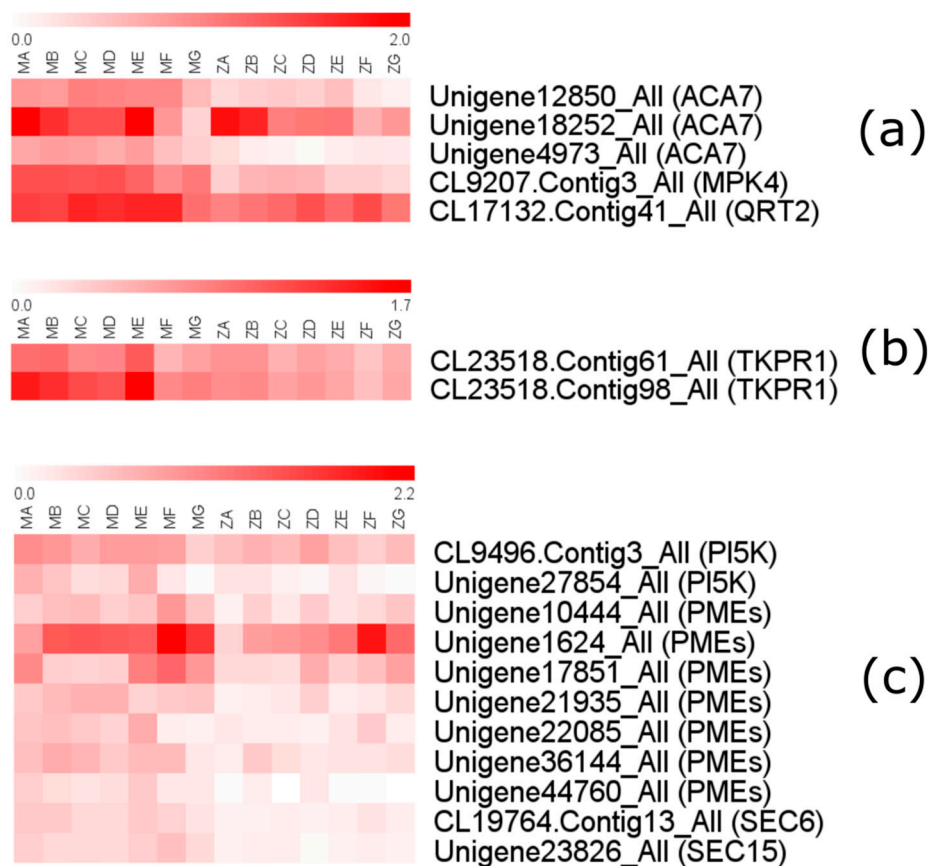


Figure 9. Cont.

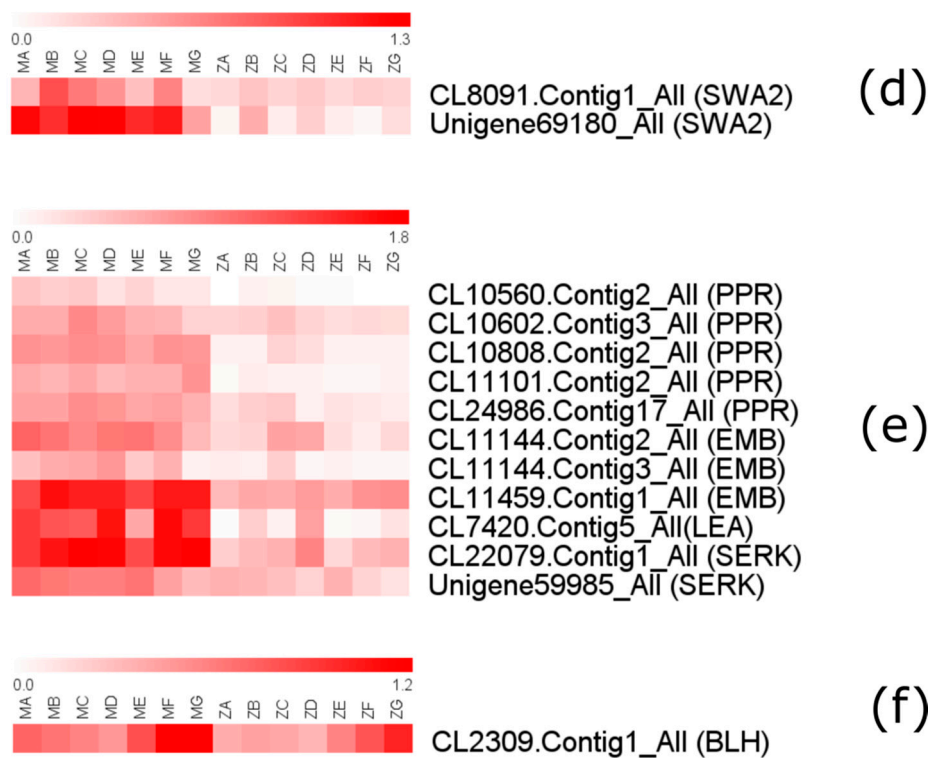


Figure 9. Expression levels of DEGs related to the reproductive process at successive pinecone developmental stages of *P. massoniana* and the Z pine. Expression level values in this figure have been transformed to a $\log_{10}(\text{FPKM}+1)$ value. (a) pollen development; (b) pollen exine formation; (c) pollen tube growth; (d) female gametophyte; (e) embryo development; (f) embryo sac.

ACA7 belongs to the auto-regulated Ca^{2+} -ATPase family, which is exclusively detected in developing flowers of *Arabidopsis* and participates in the regulation of Ca^{2+} homeostasis [28]. *MPK4* plays an important role in plant growth, development and male fertility [29]. *QRT2* is necessary for pollen grain separation and is also involved in pollen development [30]. *TKPR1* takes part in a biosynthetic pathway leading to hydroxylated α -pyrone compounds [31]. *SHT* encodes an acyltransferase that conjugates spermidine to hydroxycinnamic acids, impacting the composition of the *Arabidopsis* pollen wall [32,33]. *NPG1* in *Arabidopsis* is specifically required for pollen germination [34] and not for pollen development [35]. A type B *PI5K* mediates *Arabidopsis* and *Nicotiana* pollen tube growth by regulating apical pectin secretion [36]. *PMEs* and its pro-region adjust cell wall dynamics of growing pollen tubes in *Nicotiana tabacum* [37]. The exocyst contributes to the morphogenesis of polarized cells in many eukaryotes, for example, *SEC8* facilitates the initiation and maintenance of polarized growth of pollen tubes [38]. *SWK2* has an essential role in the coordinated mitotic progression of the female gametophyte in *Arabidopsis* [39]. Absence of *CRINKLY4* could cause an inhibition of aleurone, which is in charge of differentiation normal progression over the endosperm surface development [40]. *PPR* is required for embryo and seed viability in *Arabidopsis*, its absence leading to embryo abortion [41,42]. *EYE* controls golgi-localized proteins, that have an important role in cell and organ expansion [43]. *EDD1* encodes plastid and mitochondria, functional absence mutation of *EDD1* causes embryo lethality [44]. *LEA* and *SERK* play key roles during embryogenesis and *SERK* is essential for embryogenic competence [45,46]. Misexpression of *BLH1* leads to a cell-fate switch of synergid to egg cell in the *Arabidopsis* *eostre* mutant embryo sac [47].

4. Discussion

4.1. Gymnosperm Gene Annotation Using Transcriptome Data

Transcriptome analysis based on RNA sequencing is an effective way to explore the huge genomes of plants like gymnosperms. Several RNA sequencing studies related to gymnosperms have previously been reported [48–51]. Yet until now few studies have focused on the development of *Pinus* reproductive organs. In *Pinus tabuliformis*, unusual bisexual cones were found; here, the gene expression pattern of *MADS-box* transcription factors, *FT/TFL1-like* and *LFY/NDLY* genes was compared between unisexual and bisexual cones [52]. In *Pinus bungeana*, 39.62 Gb of RNA sequencing data was analyzed from two kinds of sexual cones, obtaining 85,305 unigenes, 53,944 (63.23%) of which were annotated in public databases [53].

In this study, we collected a total of 160 Gb of RNA sequence data from *P. massoniana* and its introgression hybrid at seven different stages of pinecone development. N50 is a key parameter in genome or transcriptome assembly. It is defined as the sequence length of the shortest contig at 50% of the entire genome or transcriptome length. In principle, the higher the N50 value, the better the sequencing quality. We obtained an N50 of 2494 bp in all-unigene, compared to previously obtained values of (N50 = 551 bp) for *Picea abies* [48] and (N50 = 1942 bp) for *P. bungeana* [53], which means the quality of our sequencing data improves on previously available data.

A total of 30,943 genes (47.05%, rank 1) were annotated to *Picea sitchensis* through the Nr database, with further annotations being 1174 (1.79%, rank 7) to *Pinus tabuliformis*, 1074 (1.63%, rank 8) to *Pinus taeda*, 697 (1.06%, rank 12) to *Pinus monticola*, 401 (0.61%, rank 18) to *Pinus radiata* and 376 (0.57%, rank 19) to *Picea abies*; all these species are conifers and belong to the Pinaceae family. Out of these, two (*Pinus taeda* and *Picea abies*) had their genomes sequenced [1,54]. The genome sizes of *Pinus taeda* and *Picea abies* are 21.6 Gb and 19.7 Gb, respectively. Pines have an estimated genome size ranging from 18 Gb to 40 Gb [55–57]. This indicated that a lot of novel genes in *P. massoniana* and the Z pine still completely unknown and expect to discover more in the future.

4.2. Impact of Introgression in Expression Levels

Introgressive hybridization implies repeated backcrossing of hybrids with parental species [58]. Hybridization between pines exists frequently in nature [59,60]. As the pollen of pines is mainly moved by wind, it could spread to a vast area. In that case, the element consisting of individuals could be with various proportions of parental genomes. Therefore, those differences between individuals could lead to diversity on gene expression, particularly in genes relate to reproduction. In sample collection, we conducted mixture of cones for each sample code and also made a mixture apply to RNA isolation to reduce the possible expression bias through analysis process.

4.3. Differential Expression of Reproductive Genes Could Relate to Delayed Maturation of Z Pinecones

Within the *Pinus* genus, some female cones take 1.5 to 3 years to mature after pollination, while for *P. massoniana* and the Z pine specifically, it takes around 1.5 years to do that. Around the middle and lower reaches of the Yangtze River, these two pines are often pollinated in April and mature cones emerge in November of the next year, a long time compared to most angiosperms. The structure of a pinecone is rather complex compared to an angiosperm flower. Therefore, more genes relate to reproduction may exist in cone than in flower and more pathways of these genes may occur in this process also.

Genes directly related to reproduction in gymnosperms have only rarely been reported. One of them is *MADS* genes, which are well studied relatively, for example in *Gnetum* spp. [61,62], *Ginkgo biloba* [63], *Picea abies* [64] and *Cryptomeria japonica* [4]. *LEAFY* is also a crucial kind of gene that involve in reproductive process in *Welwitschia mirabilis* [65] and *Pinus caribaea* var. *Caribaea* [66], as well as *NEEDLY* in *Pinus radiata* [67]. We collected several such genes from model plants (e.g., *Arabidopsis thaliana*) and analyzed their expression level in the two pines. We found that *ACA7*, *MPK4*, *QRT2*,

TKPR1, *PI5K*, *PMEs*, *SEC6*, *SEC15*, *SWK2*, *PPR*, *EMB*, *LEA*, *SERK*, *BLH1* showed a higher expression level in *P. massoniana* than in the Z pine. This result indicates that the Z pine may have a lowered expression level of genes related to pollen development, pollen exine formation, pollen tube growth and female gametophyte, embryo and/or embryo sac development, compared to *P. massoniana*. This outcome provides further understanding towards a possible molecular mechanism responsible for the altered reproduction process of the Z pine in comparison to *P. massoniana*.

5. Conclusions

P. massoniana and *P. hwangshanensis* mainly grow in southern China and produce an introgression hybrid, which we here temporarily named ‘the Z pine,’ on Mt. Lushan, where both species can be found. This Z pine has morphological characters derived from both parent species, yet has an ultra-low germination and ripening rate. In order to understand the molecular mechanism that might be causing this delayed reproduction, we collected cones from *P. massoniana* and the Z pine of seven successive developmental stages and determined their transcriptome. Herein we might discover differentially expressed genes underlying the observed reproductive delay. We obtained 93,291 unigenes with an average length size of 1987 bp and 2494 bp of N50. We identified significantly differentially expressed genes (DEGs) in all seven cone growth stages. We screened for DEGs related to reproduction, such as pollen tube growth, development of the female gametophyte and embryo and so forth. Several potentially vital genes were identified and the expression levels of the two pines were compared and analyzed. These results may offer insight into the molecular mechanisms of reproductive process between the two pines and several other plants that with similar differential mode.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1999-4907/10/3/230/s1>. Table S1: Quality of original sequencing data of *P. massoniana* and the Z pine; Table S2: Quality of clean data of *P. massoniana* and the Z pine; Table S3: Sequences of unigenes for qRT-PCR validation; Table S4: Primers for qRT-PCR; Table S5. Genes relate to reproduction process in *P. massoniana* and the Z pine; Figure S1: Length distribution of all unigene sequences; Figure S2: Venn diagram of annotation against to five databases: NCBI Nr, NCBI Nt, SwissProt, COG and KEGG for all unigenes; Figure S3. The species distribution of *P. massoniana* and the Z pine unigenes against the NCBI Nr database; Figure S4: Differentially expressed genes analysis between *P. massoniana* and the Z pine of each stage; Figure S5: Gene ontology (GO) functional classification of Z pine versus *P. massoniana* DEGs in seven stages; Figure S6: DEGs relate to reproduction in GO classification of Z pine versus *P. massoniana* in seven stages; Figure S7: KEGG pathway enrichment analysis of Z pine versus *P. massoniana* in seven stages; Figure S8: Comparison of unigene expression results between RNA sequencing (FPKM) and qRT-PCR.

Author Contributions: Conceptualization, J.X., T.Y. and J.S.; Formal analysis, J.M., Y.C., L.Y., H.H., H.Z., Z.G. and J.Y.; Funding acquisition, J.X.; Investigation, J.M., Y.C., L.Y., H.H., H.Z., Z.G. and J.Y.; Project administration, J.X.; Writing—original draft, J.M.; Writing—review & editing, J.M.

Funding: This research was funded by National Natural Science Foundation of China (31270661) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

Acknowledgments: We would like to give thanks to Youxin Du, Qiang Huang and Benzhong Zhou from Lushan Botanical Garden, Jiangxi Province and Chinese Academy of Sciences for collecting samples on the Mt. Lushan. We also thank two reviewers for insightful comments on this article. Special thanks go to editors for their help in formulating the revisions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Nystedt, B.; Street, N.R.; Wetterbom, A.; Zuccolo, A.; Lin, Y.C.; Scofield, D.G.; Vezzi, F.; Delhomme, N.; Giacomello, S.; Alexeyenko, A. The Norway spruce genome sequence and conifer genome evolution. *Nature* **2013**, *497*, 579–584. [[CrossRef](#)] [[PubMed](#)]
2. Syring, J.; Farrell, K.; Businský, R.; Cronn, R.; Liston, A. Widespread genealogical nonmonophyly in species of *Pinus subgenus* *Strobus*. *Syst. Biol.* **2007**, *56*, 163–181. [[CrossRef](#)] [[PubMed](#)]

3. Carlsbecker, A.; Sundström, J.F.; Englund, M.; Uddenberg, D.; Izquierdo, L.; Kvarnheden, A.; Vergara Silva, F.; Engström, P. Molecular control of normal and acrocona mutant seed cone development in Norway spruce (*Picea abies*) and the evolution of conifer ovule-bearing organs. *New Phytol.* **2013**, *200*, 261–275. [[CrossRef](#)] [[PubMed](#)]
4. Katahata, S.; Futamura, N.; Igasaki, T.; Shinohara, K. Functional analysis of *SOC1-like* and *AGL6-like* *MADS-box* genes of the gymnosperm *Cryptomeria japonica*. *Tree Genet. Genomes* **2014**, *10*, 317–327. [[CrossRef](#)]
5. Uddenberg, D.; Reimegård, J.; Clapham, D.; Almqvist, C.; von Arnold, S.; Emanuelsson, O.; Sundström, J.F. Early cone setting in *Picea abies* acrocona is associated with increased transcriptional activity of a *MADS* box transcription factor. *Plant Physiol.* **2013**, *161*, 813–823. [[CrossRef](#)] [[PubMed](#)]
6. Dubos, C.; Stracke, R.; Grotewold, E.; Weisshaar, B.; Martin, C.; Lepiniec, L. *MYB* transcription factors in *Arabidopsis*. *Trends Plant Sci.* **2010**, *15*, 573–581. [[CrossRef](#)] [[PubMed](#)]
7. Hennig, L.; Taranto, P.; Walser, M.; Schönrock, N.; Gruissem, W. *Arabidopsis MSII* is required for epigenetic maintenance of reproductive development. *Development* **2003**, *130*, 2555–2565. [[CrossRef](#)] [[PubMed](#)]
8. Zhai, D.; He, Z.; Feng, J.; Zheng, Y. Study on introgression between *Pinus hwangshanensis* and *Pinus massoniana* by using inter-simple sequence repeat marker (ISSR). *For. Sci. Technol.* **2012**, *37*, 4–6. (In Chinese)
9. Luo, S.; Zou, H.; Liang, S. Study on the introgressive hybridization between *Pinus hwangshanensis* and *P. massoniana*. *Sci. Silvae Sin.* **2001**, *37*, 118–122. (In Chinese)
10. Li, S.; Chen, Y.; Gao, H.; Yin, T. Potential chromosomal introgression barriers revealed by linkage analysis in a hybrid of *Pinus massoniana* and *P. hwangshanensis*. *BMC Plant Biol.* **2010**, *10*, 37. [[CrossRef](#)] [[PubMed](#)]
11. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)] [[PubMed](#)]
12. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512. [[CrossRef](#)] [[PubMed](#)]
13. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Xian, A.; Fan, L.; Raychowdhury, R.; Zeng, Q. Trinity: Reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* **2011**, *29*, 644–652. [[CrossRef](#)] [[PubMed](#)]
14. Perteau, G.; Huang, X.; Liang, F.; Antonescu, V.; Sultana, R.; Karamycheva, S.; Lee, Y.; White, J.; Cheung, F.; Parvizi, B. TIGR Gene Indices clustering tools (TGICL): A software system for fast clustering of large EST datasets. *Bioinformatics* **2003**, *19*, 651–652. [[CrossRef](#)] [[PubMed](#)]
15. Schneider, M.; Lane, L.; Boutet, E.; Lieberherr, D.; Tognolli, M.; Bougueleret, L.; Bairoch, A. The UniProtKB/Swiss-Prot knowledgebase and its Plant Proteome Annotation Program. *J. Proteom.* **2009**, *72*, 567–573. [[CrossRef](#)] [[PubMed](#)]
16. Pruitt, K.D.; Tatusova, T.; Maglott, D.R. NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **2005**, *33*, 501–504. [[CrossRef](#)] [[PubMed](#)]
17. Kanehisa, M.; Araki, M.; Goto, S.; Hattori, M.; Hirakawa, M.; Itoh, M.; Katayama, T.; Kawashima, S.; Okuda, S.; Tokimatsu, T. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **2008**, *36*, 480–484. [[CrossRef](#)] [[PubMed](#)]
18. Ogata, H.; Goto, S.; Sato, K.; Fujibuchi, W.; Bono, H.; Kanehisa, M. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **1999**, *27*, 29–34. [[CrossRef](#)] [[PubMed](#)]
19. Tatusov, R.L.; Fedorova, N.D.; Jackson, J.D.; Jacobs, A.R.; Kiryutin, B.; Koonin, E.V.; Krylov, D.M.; Mazumder, R.; Mekhedov, S.L.; Nikolskaya, A.N.; et al. The COG database: An updated version includes eukaryotes. *BMC Bioinf.* **2003**, *4*, 41. [[CrossRef](#)] [[PubMed](#)]
20. Young, M.D.; Wakefield, M.J.; Smyth, G.K.; Oshlack, A. Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biol.* **2010**, *11*, R14. [[CrossRef](#)] [[PubMed](#)]
21. Conesa, A.; Terol, J.; Robles, M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674–3676. [[CrossRef](#)] [[PubMed](#)]
22. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* **2011**, *12*, 323. [[CrossRef](#)] [[PubMed](#)]
23. Trapnell, C.; Williams, B.A.; Pertea, G.; Mortazavi, A.; Kwan, G.; Van Baren, M.J.; Salzberg, S.L.; Wold, B.J.; Pachter, L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **2010**, *28*, 511. [[CrossRef](#)] [[PubMed](#)]

24. Anders, S. Analysing RNA-Seq data with the DESeq package. *Mol. Biol.* **2010**, *43*, 1–17.
25. Ernst, J.; Bar-Joseph, Z. STEM: A tool for the analysis of short time series gene expression data. *BMC Bioinf.* **2006**, *7*, 191. [[CrossRef](#)] [[PubMed](#)]
26. Ye, J.; Coulouris, G.; Zaretskaya, I.; Cutcutache, I.; Rozen, S.; Madden, T.L. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinf.* **2012**, *13*, 134. [[CrossRef](#)] [[PubMed](#)]
27. Livak, K.J.; Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **2001**, *25*, 402–408. [[CrossRef](#)] [[PubMed](#)]
28. Lucca, N.; León, G. *Arabidopsis* ACA7, encoding a putative auto-regulated Ca^{2+} -ATPase, is required for normal pollen development. *Plant Cell Rep.* **2012**, *31*, 651–659. [[CrossRef](#)] [[PubMed](#)]
29. Zeng, Q.; Chen, J.G.; Ellis, B.E. *AtMPK4* is required for male-specific meiotic cytokinesis in *Arabidopsis*. *Plant J.* **2011**, *67*, 895–906. [[CrossRef](#)] [[PubMed](#)]
30. Ogawa, M.; Kay, P.; Wilson, S.; Swain, S.M. *Arabidopsis* dehiscence zone polygalacturonase 1 (*ADPG1*), *ADPG2*, and *QUARTET2* are polygalacturonases required for cell separation during reproductive development in *Arabidopsis*. *Plant Cell* **2009**, *21*, 216–233. [[CrossRef](#)] [[PubMed](#)]
31. Grienenberger, E.; Kim, S.S.; Lallemand, B.; Geoffroy, P.; Heintz, D.; de Azevedo Souza, C.; Heitz, T.; Douglas, C.J.; Legrand, M. Analysis of TETRAKETIDE α -PYRONE REDUCTASE function in *Arabidopsis thaliana* reveals a previously unknown, but conserved, biochemical pathway in sporopollenin monomer biosynthesis. *Plant Cell* **2010**, *110*. [[CrossRef](#)]
32. Grienenberger, E.; Besseau, S.; Geoffroy, P.; Debayle, D.; Heintz, D.; Lapierre, C.; Pollet, B.; Heitz, T.; Legrand, M. A *BAHD* acyltransferase is expressed in the tapetum of *Arabidopsis* anthers and is involved in the synthesis of hydroxycinnamoyl spermidines. *Plant J.* **2009**, *58*, 246–259. [[CrossRef](#)] [[PubMed](#)]
33. Dobritsa, A.A.; Geanconteri, A.; Shrestha, J.; Carlson, A.; Kooyers, N.; Coerper, D.; Urbanczyk-Wochniak, E.; Bench, B.J.; Sumner, L.W.; Swanson, R. A large-scale genetic screen in *Arabidopsis thaliana* to identify genes involved in pollen exine production. *Plant Physiol.* **2011**, *157*, 947–970. [[CrossRef](#)] [[PubMed](#)]
34. Shin, S.; Golovkin, M.; Reddy, A.S. A pollen-specific calmodulin-binding protein, *NPG1*, interacts with putative pectate lyases. *Sci. Rep.* **2014**, *4*, 5263. [[CrossRef](#)] [[PubMed](#)]
35. Golovkin, M.; Reddy, A.S. A calmodulin-binding protein from *Arabidopsis* has an essential role in pollen germination. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 10558–10563. [[CrossRef](#)] [[PubMed](#)]
36. Ischebeck, T.; Stenzel, I.; Heilmann, I. Type B phosphatidylinositol-4-phosphate 5-kinases mediate *Arabidopsis* and *Nicotiana tabacum* pollen tube growth by regulating apical pectin secretion. *Plant Cell* **2008**, *20*, 3312–3330. [[CrossRef](#)] [[PubMed](#)]
37. Bosch, M.; Cheung, A.Y.; Hepler, P.K. Pectin methylesterase, a regulator of pollen tube growth. *Plant Physiol.* **2005**, *138*, 1334–1346. [[CrossRef](#)] [[PubMed](#)]
38. Cole, R.A.; Synek, L.; Zarsky, V.; Fowler, J.E. *SEC8*, a subunit of the putative *Arabidopsis* exocyst complex, facilitates pollen germination and competitive pollen tube growth. *Plant Physiol.* **2005**, *138*. [[CrossRef](#)] [[PubMed](#)]
39. Li, N.; Yuan, L.; Liu, N.; Shi, D.; Li, X.; Tang, Z.; Liu, J.; Sundaresan, V.; Yang, W. *SLOW WALKER2*, a *NOC1/MAK21* homologue, is essential for coordinated cell cycle progression during female gametophyte development in *Arabidopsis*. *Plant Physiol.* **2009**, *151*, 1486–1497. [[CrossRef](#)] [[PubMed](#)]
40. Becraft, P.W.; Stinard, P.S.; McCarty, D.R. *CRINKLY4*: A *TNFR-like* receptor kinase involved in maize epidermal differentiation. *Science* **1996**, *273*, 1406–1409. [[CrossRef](#)] [[PubMed](#)]
41. Lurin, C.; Andrés, C.; Aubourg, S.; Bellaoui, M.; Bitton, F.; Bruyère, C.; Caboche, M.; Debast, C.; Gualberto, J.; Hoffmann, B.; et al. Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *Plant Cell* **2004**, *16*, 2089–2103. [[CrossRef](#)] [[PubMed](#)]
42. Cushing, D.A.; Forsthoefel, N.R.; Gestaut, D.R.; Vernon, D.M. *Arabidopsis emb175* and other *ppr* knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis. *Planta* **2005**, *221*, 424–436. [[CrossRef](#)] [[PubMed](#)]
43. Ishikawa, T.; Machida, C.; Yoshioka, Y.; Ueda, T.; Nakano, A.; Machida, Y. *EMBRYO YELLOW* gene, encoding a subunit of the conserved oligomeric Golgi complex, is required for appropriate cell expansion and meristem organization in *Arabidopsis thaliana*. *Genes Cells* **2008**, *13*, 521–535. [[CrossRef](#)] [[PubMed](#)]
44. Berg, M.; Rogers, R.; Muralla, R.; Meinke, D. Requirement of aminoacyl-tRNA synthetases for gametogenesis and embryo development in *Arabidopsis*. *Plant J.* **2005**, *44*, 866–878. [[CrossRef](#)] [[PubMed](#)]

45. Galau, G.A.; Bijaisoradat, N.; Hughes, D.W. Accumulation kinetics of cotton late embryogenesis-abundant mRNAs and storage protein mRNAs: Coordinate regulation during embryogenesis and the role of abscisic acid. *Dev. Biol.* **1987**, *123*, 198–212. [[CrossRef](#)]
46. Hecht, V.; Vielle-Calzada, J.; Hartog, M.V.; Schmidt, E.D.; Boutilier, K.; Grossniklaus, U.; de Vries, S.C. The *Arabidopsis* SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1 gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture. *Plant Physiol.* **2001**, *127*, 803–816. [[CrossRef](#)] [[PubMed](#)]
47. Pagnussat, G.C.; Yu, H.; Sundaresan, V. Cell-fate switch of synergid to egg cell in *Arabidopsis* eostre mutant embryo sacs arises from misexpression of the *BEL1*-like homeodomain gene *BLH1*. *Plant Cell* **2007**, *19*, 3578–3592. [[CrossRef](#)] [[PubMed](#)]
48. Chen, J.; Uebbing, S.; Gyllenstrand, N.; Lagercrantz, U.; Lascoux, M.; Källman, T. Sequencing of the needle transcriptome from Norway spruce (*Picea abies* Karst L.) reveals lower substitution rates, but similar selective constraints in gymnosperms and angiosperms. *BMC Genom.* **2012**, *13*, 589. [[CrossRef](#)] [[PubMed](#)]
49. Liu, J.J.; Sturrock, R.N.; Benton, R. Transcriptome analysis of *Pinus monticola* primary needles by RNA-seq provides novel insight into host resistance to *Cronartium ribicola*. *BMC Genom.* **2013**, *14*, 884. [[CrossRef](#)] [[PubMed](#)]
50. Cañas, R.A.; Feito, I.; Fuentemaqueda, J.F.; Ávila, C.; Majada, J.; Cánovas, F.M. Transcriptome-wide analysis supports environmental adaptations of two *Pinus pinaster* populations from contrasting habitats. *BMC Genom.* **2015**, *16*, 909. [[CrossRef](#)] [[PubMed](#)]
51. Wu, Y.; Guo, J.; Zhou, Q.; Xin, Y.; Wang, G.; Xu, L.A. *De novo* transcriptome analysis revealed genes involved in flavonoid biosynthesis, transport and regulation in *Ginkgo biloba*. *Ind. Crops Prod.* **2018**, *124*, 226–235. [[CrossRef](#)]
52. Niu, S.; Yuan, H.; Sun, X.; Porth, I.; Li, Y.; El-Kassaby, Y.A.; Li, W. A transcriptomics investigation into pine reproductive organ development. *New Phytol.* **2016**, *209*, 1278–1289. [[CrossRef](#)] [[PubMed](#)]
53. Duan, D.; Jia, Y.; Yang, J.; Li, Z.H. Comparative Transcriptome Analysis of Male and Female Conelets and Development of Microsatellite Markers in *Pinus bungeana*, an Endemic Conifer in China. *Genes* **2017**, *8*, 393. [[CrossRef](#)] [[PubMed](#)]
54. Neale, D.B.; Wegrzyn, J.L.; Stevens, K.A.; Zimin, A.V.; Puiu, D.; Crepeau, M.W.; Cardeno, C.; Koriabine, M.; Holtz-Morris, A.E.; Liechty, J.D.; et al. Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biol.* **2014**, *15*, R59. [[CrossRef](#)] [[PubMed](#)]
55. Grotkopp, E.; Rejmánek, M.; Sanderson, M.J.; Rost, T.L. Evolution of genome size in pines (*Pinus*) and its life-history correlates: Supertree analyses. *Evolution* **2004**, *58*, 1705–1729. [[CrossRef](#)] [[PubMed](#)]
56. Wakamiya, I.; Newton, R.J.; Johnston, J.S.; Price, H.J. Genome size and environmental factors in the genus *Pinus*. *Am. J. Bot.* **1993**, *80*, 1235–1241. [[CrossRef](#)]
57. Ohri, D.; Khoshoo, T.N. Genome Size in Gymnosperms. *Plant Syst. Evol.* **1986**, *153*, 119–132. [[CrossRef](#)]
58. Anderson, E.; Hubricht, L. Hybridization in *Tradescantia*. III. The evidence for introgressive hybridization. *Am. J. Bot.* **1938**, *25*, 396–402. [[CrossRef](#)]
59. Wang, B.; Mao, J.F.; Gao, J.; Zhao, W.E.; Wang, X.R. Colonization of the Tibetan Plateau by the homoploid hybrid pine *Pinus densata*. *Mol. Ecol.* **2011**, *20*, 3796–3811. [[CrossRef](#)] [[PubMed](#)]
60. Watano, Y.; Kanai, A.; Tani, N. Genetic structure of hybrid zones between *Pinus pumila* and *P. parviflora* var. *pentaphylla* (Pinaceae) revealed by molecular hybrid index analysis. *Am. J. Bot.* **2004**, *91*, 65–72. [[CrossRef](#)] [[PubMed](#)]
61. Shindo, S.; Ito, M.; Ueda, K.; Kato, M.; Hasebe, M. Characterization of *MADS* genes in the gymnosperm *Gnetum parvifolium* and its implication on the evolution of reproductive organs in seed plants. *Evol. Dev.* **1999**, *1*, 180–190. [[CrossRef](#)] [[PubMed](#)]
62. Becker, A.; Saedler, H.; Theissen, G. Distinct *MADS-box* gene expression patterns in the reproductive cones of the gymnosperm *Gnetum gnemon*. *Dev. Genes Evol.* **2003**, *213*, 567–572. [[CrossRef](#)] [[PubMed](#)]
63. Jager, M.; Hassanin, A.; Manuel, M.; Le Guyader, H.; Deutsch, J. *MADS-box* genes in *Ginkgo biloba* and the evolution of the *AGAMOUS* family. *Mol. Biol. Evol.* **2003**, *20*, 842–854. [[CrossRef](#)] [[PubMed](#)]
64. Carlsbecker, A.; Tandre, K.; Johanson, U.; Englund, M.; Engström, P. The *MADS-box* gene *DAL1* is a potential mediator of the juvenile-to-adult transition in Norway spruce (*Picea abies*). *Plant J.* **2004**, *40*, 546–557. [[CrossRef](#)] [[PubMed](#)]

65. Moyroud, E.; Monniaux, M.; Thévenon, E.; Dumas, R.; Scutt, C.; Frohlich, M.; Parcy, F. A link between *LEAFY* and *B*-gene homologues in *Welwitschia mirabilis* sheds light on ancestral mechanisms prefiguring floral development. *New Phytol.* **2017**, *216*, 469–481. [[CrossRef](#)] [[PubMed](#)]
66. Dornelas, M.; Rodriguez, A. A *FLORICAULA/LEAFY* gene homolog is preferentially expressed in developing female cones of the tropical pine *Pinus caribaea* var. *caribaea*. *Genet. Mol. Biol.* **2005**, *28*, 299–307. [[CrossRef](#)]
67. Mouradov, A.; Glassick, T.; Hamdorf, B.; Murphy, L.; Fowler, B.; Marla, S.; Teasdale, R. *NEEDLY*, a *Pinus radiata* ortholog of *FLORICAULA/LEAFY* genes, expressed in both reproductive and vegetative meristems. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6537–6542. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).