Tree Genetics and Molecular Breeding 2022, Vol.12, No.7, 1-14
http://genbreedpublisher.com/index.php/tgmb

Research Article Open Access

Transcriptome Sequencing and Bioinformatic Analysis of *Eucalyptus cloeziana* Terminal Buds

Jun Lan [1], Weixin Jiang [2], Lei Zhang [1], Xinyuan Liang [2], Tiandao Bai [2]
1 Dongmen Forest Farm of Guangxi Zhuang Autonomous Region, Chongzuo, 532108, China
2 Forestry College of Guangxi University, Nanning, 530000, China
Corresponding author email: btdman20@163.com
Tree Genetics and Molecular Breeding, 2022, Vol.12, No.7 doi: 10.5376/tgmb.2022.12.0007
Received: 13 Jun., 2022
Accepted: 20 Jun., 2022
Published: 25 Jun., 2022

Preferred citation for this article:
Lan J., Jiang W.X., Zhang L., Liang X.Y., and Bai T.D., 2022, Transcriptome sequencing and bioinformatic analysis of *Eucalyptus cloeziana* terminal buds, Tree Genetics and Molecular Breeding, 12(7): 1-14 (doi: 10.5376/tgmb.2022.12.0007)

Abstract In order to obtain transcriptome data and predict the key gene function of *Eucalyptus cloeziana*, an important timber species in China, the Illumina HiSeq X Ten sequencing technology was conducted to carry out transcriptome sequencing of *E. cloeziana* terminal buds. After Trinity assembly and splicing, Blast software was used to compare and annotate the high-quality Unigene with seven public databases including NR, Swiss-prot, KOG, Go, KEGG, etc. SSR sites search and analysis was performed by MISA software. A total of 26 587 high quality Unigenes were obtained from the terminal bud of *E. cloeziana* with an average length of 1279.69 bp. A total of 22 099 Unigenes were successfully annotated in at least one biological database. Of these, 11 507 Unigenes were successfully annotated with 25 functions in KOG database, and the most common function was general functional genes prediction. In the GO database, the 14 105 Unigenes annotated were matched to 50 functional gene groups in 3 categories of biological function, cell component and molecular function, respectively. Among them, the biological processes accounted for the largest proportion. Through KEGG pathways analysis, 7 117 Unigenes were successfully annotated and 127 metabolic pathways were detected, with the most abundant metabolism-related genes. Moreover, 1 021 Unigenes annotations were assigned with 65 families of transcription factor (TF) database, among which the bHLH and MYB had the largest proportion. A total of 3 274 Unigenes were annotated in PRG database and formed into 13 resistant gene categories, among which RLP and TNL had the largest number of matched genes. In addition, 12 366 SSR sites were detected by MISA software, with a distribution density of 1/2.75 kb and a variety of repeating primitive types. In this study, abundant transcriptome information from terminal bud was obtained by using high-throughput sequencing, which was beneficial for molecular assisted breeding of *E. cloeziana*.
Keywords *Eucalyptus cloeziana*; Illumina HiSeq X Ten; Transcriptome; Gene annotation; Terminal buds

Eucalyptus cloeziana is a large tree of *Eucalyptus* genus in the family of Myrtaceae, also known as Kunshilan'an. The tree height can reach 55 m, the natural distribution area is located in the central and northern areas of Queensland, Australia (44°44' ~ 152°52' E, 15°45' ~ 26°41' S), and the altitude ranges 25~950 m (Xiang et al., 2008; Wang et al., 2016). Relevant research showed that *Eucalyptus cloeziana*, with straight shape, high wood hardness, high strength, insect resistance, beautiful wood patterns and relatively uniform texture structure, is an excellent material for making high-grade furniture and solid wood flooring. The wood utilization rate of *Eucalyptus cloeziana* is more than 70%, and its economic value is much higher than that of short-rotation "Sufeng'an". It is one of the medium and large diameter timber species with great cultivation value. It has been included in the list of national precious timber species and has very broad development prospects (Huang et al., 2018; Li et al., 2012a). Since 1972, *Eucalyptus cloeziana* has been introduced into China. In the 1980s, systematic studies such as introduction test, provenance selection, family test, correlation degree of wood properties and growth properties have been carried out in Guangxi, Guangdong, Hainan, Fujian, Hunan and Sichuan (Huang et al., 2018). In general, *Eucalyptus cloeziana* has strong growth potential and material characteristics with significant volume growth, for example, the average basic density of 18-year-old *Eucalyptus cloeziana* is 0.706 g/cm, and the variation range between provenances is 0.617~0.753 g/cm (Li et al., 2012b; Xiang et al., 2008); The

heritability (H^2) of different provenances of 25-year-old *Eucalyptus cloeziana* is 0.634~0.895, and the heritability (h^2) of single plant is 0.136~0.342 (Wang et al., 2016).

Transcriptome is the sum of all mRNA transcribed by a specific cell or tissue at a certain developmental stage or functional state. With the development of a new generation of high-throughput sequencing technology, the generation of a large number of transcriptome sequencing data can provide rich sequence resources for animal and plant gene expression and function analysis, molecular marker development and related metabolic pathway research (Vaattovaara et al., 2019). The development of molecular biology of *Eucalyptus cloeziana* is slow, and the lack of genomic and transcriptome information limits the research on genetic improvement of *Eucalyptus cloeziana*. At present, only Zhu et al. (2018) annotated and analyzed the gene function of the transcriptome sequence of *Eucalyptus cloeziana* root. The existing SSR markers of *Eucalyptus cloeziana* all come from the reported near source *Eucalyptus*. For example, Lü et al. (2020) used 12 SSR loci to study the genetic variation level of free pollinated progeny forest of different provenances of *Eucalyptus cloeziana*; Deng et al. (2019) analyzed the genetic variation of *Eucalyptus cloeziana* population by using the SSR loci developed in near source *Eucalyptus*, but there is no report on the development of SSR markers for *Eucalyptus cloeziana* at present. The existing phylogenetic analysis of *Eucalyptus* shows that (Steane et al., 2011), *Eucalyptus cloeziana* is far from other *eucalyptus* trees in the same genus and belongs to an independent subgenus. Therefore, the research on the genome and transcriptome of this species will help to further explore its species characteristics.

In this study, the 2-year-old *Eucalyptus cloeziana* terminal buds were used as the material, the transcriptome sequence data were obtained by Illumina HiSeq X Ten high-throughput sequencing technology, and the gene function annotation and SSR locus retrieval and analysis were carried out for all Unigenes, so as to provide a theoretical basis for the functional gene mining, genetic map construction and SSR marker development of important traits such as growth and wood properties of *Eucalyptus cloeziana*.

1 Results and Analysis

1.1 Transcriptome sequencing analysis of *Eucalyptus cloeziana*

56 157 538 Raw reads were obtained by sequencing the transcriptome of *Eucalyptus cloeziana* terminal buds, and the number of bases was 8 423 630 700 bp (8.4 Gb). After removing the low-quality original sequence, 55 203 802 clean reads were obtained, accounting for 98.3% of the original sequence. The amount of clean bases was 7 473 648 753 bp (7.5 Gb), the percentage of bases with quality of Q30 (the correct recognition rate of bases up to 99.9%) was more than 93.99%, and the content of GC was 49.97%. The above data showed that the transcriptional data of *Eucalyptus cloeziana* obtained by sequencing on Illumina hiseq platform was of good quality. Using Trinity software for assembly and splicing, a total of 26 587 Unigenes were obtained, with a total length of 34 023 022 bp, N50 of 1 851 bp, the longest and shortest sequences of 15 778 bp and 301 bp respectively, and the average length of 1 279.69 bp. There were 13 527 Unigenes between 301 and 1 000 bp, accounting for half of the total (50.87%); There were 7 836 Unigenes between 1001 and 2000 bp, accounting for 29.47% of the total; The length >2000 bp accounted for 19.65% (5 224 Unigenes) (Figure 1). In general, the transcriptome sequencing of *Eucalyptus cloeziana* terminal buds was of good quality, the sequence assembly was long, and the length range was relatively uniform.

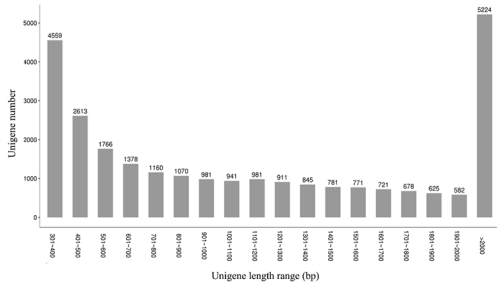

Figure 1 Unigene length distribution of *Eucalyptus cloeziana*

1.2 Basic function annotations of transcriptome in *Eucalyptus cloeziana* terminal buds

All UniGenes (26 587) of the transcriptome in *Eucalyptus cloeziana* terminal buds obtained by sequencing were compared with seven related databases (NR, COG/KOG, GO, Swiss-Prot, eggNOG, KEGG and Pfam) by BlastX for annotation, gene function and metabolic pathway analysis (Figure 2). Finally, 22 099 Unigenes (83.12%) with annotation information were obtained, of which 4 777 (17.96%) were successfully annotated in all databases. The number of Unigenes annotated in NR database was the largest, with a total of 21 969 (82.63%); The number of Unigenes annotated in KEGG database was the least, with only 7 117 (26.77%). The number of Unigenes annotated in eggNOG, Swiss-Prot, Pfam, GO and KOG databases was 19 671 (73.99%), 15 469 (58.86%), 15 208 (57.20%), 14 105 (53.05%) and 11 507 (43.38%) respectively. The Unigenes of transcriptome in *Eucalyptus cloeziana* terminal buds obtained higher annotation rate.

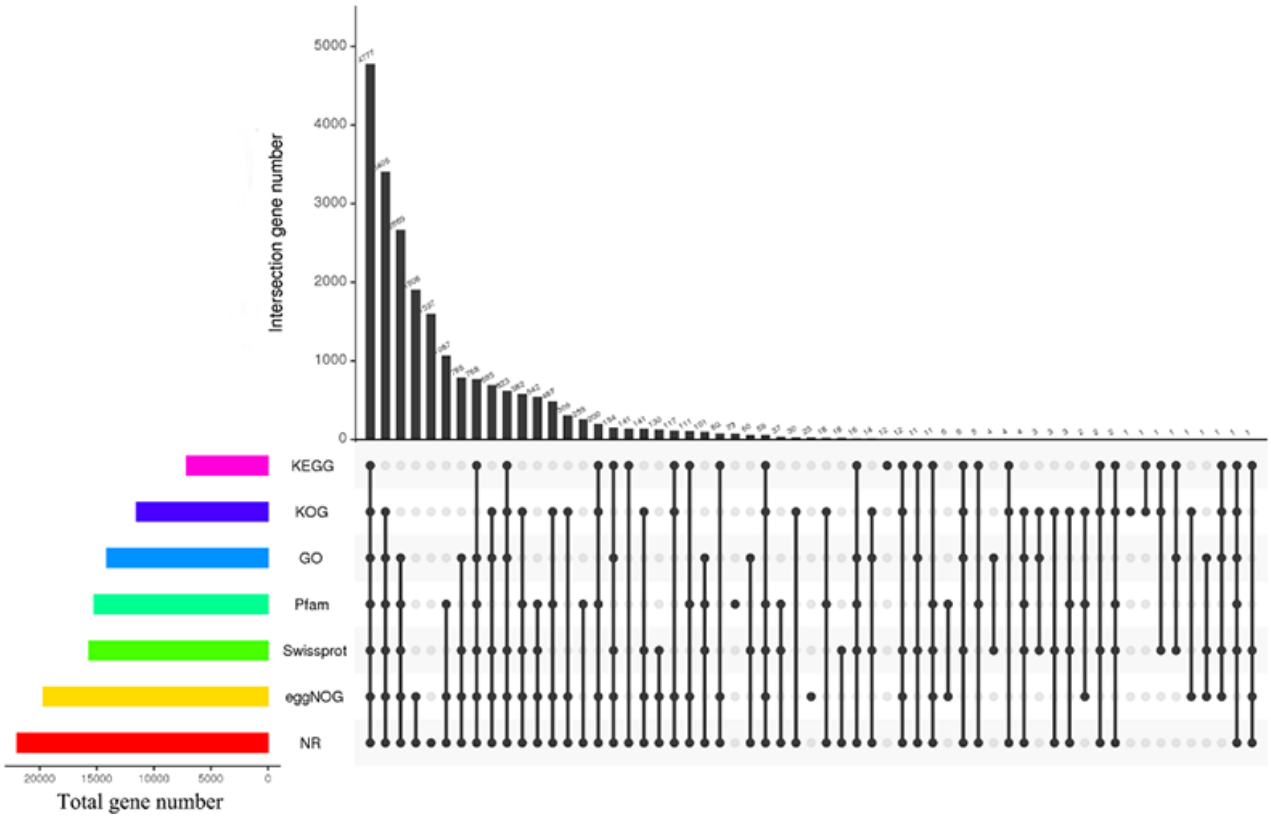

Figure 2 Venn diagram of seven databases annotations in *Eucalyptus cloeziana* Unigenes

Note: The numbers on the top bar chart represent the results of the intersection of the databases with black dots in the matrix below; The columns on the left represent the number of genes annotated to each database

1.2.1 NR function class

21 969 Unigenes of *Eucalyptus cloeziana* transcriptome were aligned with similar sequences in NR database. From the species distribution of homologous sequence (Figure 3), the matching similarity between *Eucalyptus cloeziana* and *Eucalyptus grandis* was high, and the number of Unigenes annotated to *Eucalyptus grandis* was as high as 20 767 (94.55% of the total Unigenes annotated in NR); A total of 1% of UniGenes was annotated to *Punica granatum* (86), *Vitis vinifera* (26), *Arabidopsis thaliana* (24), *Theobroma cacao* (22 articles), *Quercus suber* (22s) and *Cephalotus follicularis* (21); The remaining 2.73% (599) annotated to other species. The results reflected that *Eucalyptus grandis* had rich transcriptome data in NR database. *Eucalyptus grandis* and *Eucalyptus cloeziana* were of the same genus (*Eucalyptus*), so the number of Unigenes annotated to *Eucalyptus grandis* was the largest.

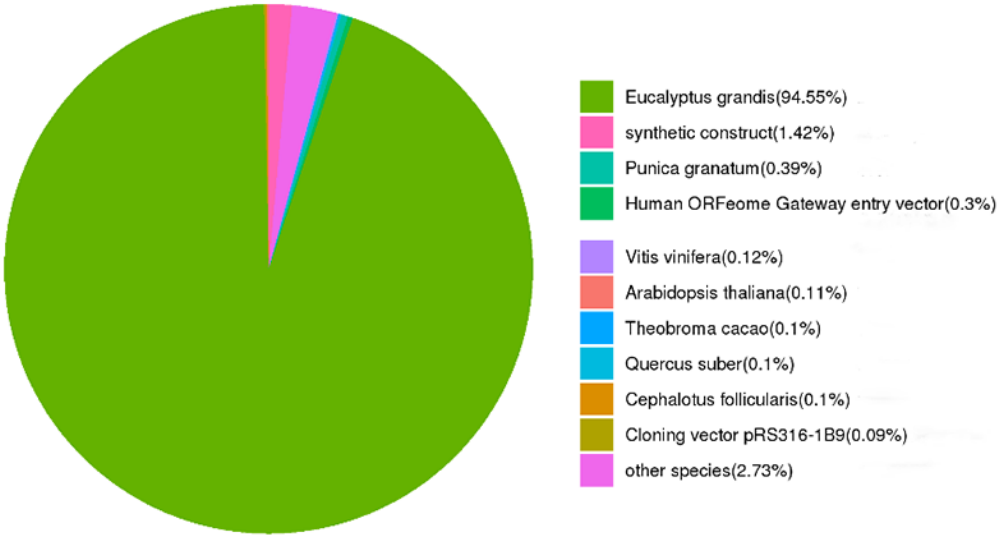

Figure 3 Species distribution of *Eucalyptus cloeziana* Unigenes in NR database

1.2.2 KOG function class

A total of 11 507 Unigenes were annotated into 25 KOG function classes, with a total number of 12 920. There were differences in the Unigene number in different function classes (Figure 4). The number of Unigenes annotated by general function prediction only (R) was the largest, including 2 399 Unigenes, accounting for 18.56% of the total number of annotations; Followed by post-translational modification, protein turnover, chaperones (O) and signal transduction mechanisms (T), the number of annotations were 1 382 (10.69%) and 1 239 (9.59%) respectively; Transcription (K), carbohydrate transport and metabolism (G), translation, ribosomal structure and biogenesis (J) and intracellular trafficking, secretion, and vesicular transport (U) were also more, with 717 (5.55%), 690 (5.34%), 595 (4.61%) and 585 (4.53%) genes respectively. Among all functional classes, nuclear structure (Y) (41), extracellular structures (W) (38) and cell motility (N) (6) was the least. The transcriptome of *Eucalyptus cloeziana* terminal buds was consistent with the annotation results obtained in most plants, and the most genes were enriched with general function and protein post-translational modification function.

1.2.3 GO function class

The GO functional annotation classification of *Eucalyptus cloeziana* terminal buds was carried out (Figure 5). A total of 14 105 Unigenes were annotated into the GO database, accounting for 53.05% of the total Unigenes, which were respectively matched to 50 GO function categories of biological pathways (processes), cellular components and molecular functions. Further analysis found that the total number of genes performing biological processes in go database was 49 872, which was divided into 22 terms. Among them, the number involved in cellular process (9400) and metabolic process (7565) was the largest, followed by response to stimulus (4623) and biological regulation (4457), and the expression of biological adhesion and cell killing was the lowest, only 11 and 10; The total number of gene annotations of cell components was 51 804, which was divided into 13 terms. Among them, the expression of cell (11 853) and cell part (11 835) was the highest, followed by organelle (9 220), and the proportion of nucleoid (45) was the lowest; The total number of annotated genes performing molecular functions is 17188, which is divided into 15 functional regions. The most representative genes are those related to binding (8 238) and catalytic activity (6 721), followed by transporter activity (1 085). The proportion of receptor regulator activity was the lowest, only 9 sequences. The number of genes involved in cellular and metabolic processes was the largest, which was consistent with expectations, reflecting the rapid growth of young *Eucalyptus cloeziana*, vigorous cell division and active metabolism in the meristem region at the top of the stem tip.

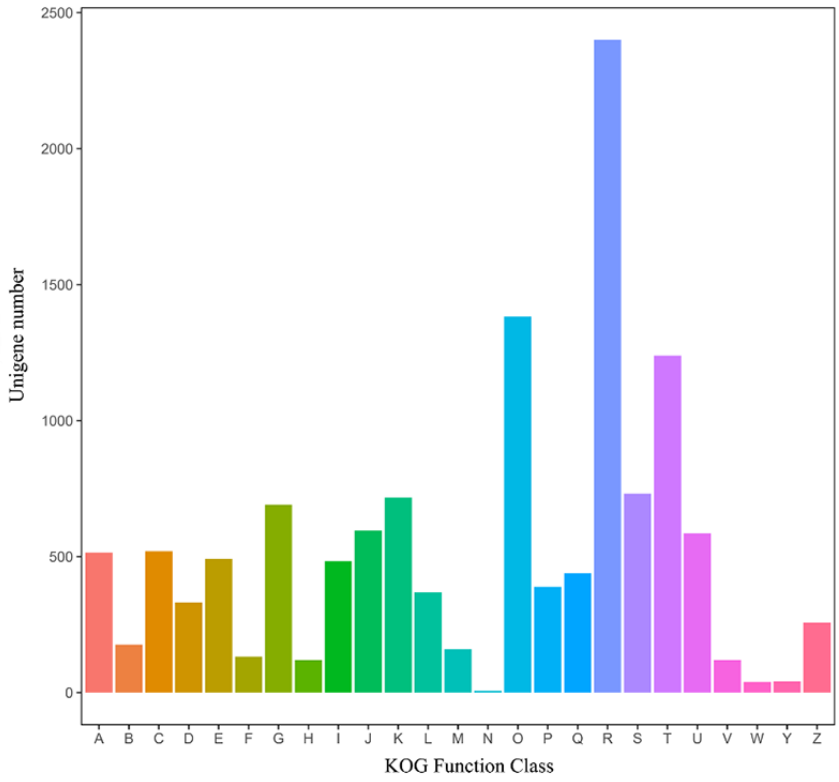

Figure 4 KOG annotation of *Eucalyptus cloeziana* Unigenes

Note: A: RNA processing and modification; B: Chromatin structure and dynamics; C: Energy production and conversion; D: Cell cycle control, cell division, chromosome partitioning; E: Amino acid transport and metabolism; F: Nucleotide transport and metabolism; G: Carbohydrate transport and metabolism; H: Coenzyme transport and metabolism; I: Lipid transport and metabolism; J: Translation, ribosomal structure and biogenesis; K: Transcription; L: Replication, recombination and repair; M: Cell wall\membrane\envelope biogenesis; N: Cell motility; O: Post-translational modification, protein turnover, chaperones; P: Inorganic ion transport and metabolism; Q: Secondary metabolites biosynthesis, transport and catabolism; R: General function prediction only; S: Function unknown; T: Signal transduction mechanisms; U: Intracellular trafficking, secretion, and vesicular transport; V: Defense mechanisms; W: Extracellular structures; Y: Nuclear structure; Z: Cytoskeleton

1.2.4 KEGG metabolic pathway

After KEGG functional annotation of *Eucalyptus cloeziana* transcriptome UniGenes, a total of 7 117 Unigenes were found to be involved in 127 metabolic pathways (Figure 6). According to their functions, these successfully annotated Unigenes were divided into five categories: metabolism, genetic information processing, cellular processes, environment information processing and organismal systems, and the number was 2 775, 1 578, 315, 278 and 195 respectively. The five metabolic pathways were further divided into 18 subclasses, and the metabolic pathways related to metabolism was the most abundant, with 10 in total. Among them, the genes involved in carbohydrate metabolism (660, 16.03%) accounted for the largest proportion, followed by amino acid metabolism (409, 9.93%) and energy metabolism (353, 8.57%); There are 4 metabolic pathways related to genetic information processing, among which the genes involved in translation (642, 15.59%) accounted for the largest proportion, followed by folding, sorting and degradation (526, 12.77%); The other pathways mainly involved transport and catabolism (315, 7.65%), signal transduction (253, 6.14%) and environmental adaptation (195, 4.74%). The proportion of genes involved in metabolic pathways related to carbohydrate metabolism in *Eucalyptus cloeziana* terminal buds transcriptome was the largest, which was consistent with the gene annotation results of more metabolic processes obtained by GO function, which can provide bioinformatics basis for functional gene mining and metabolomics of important traits such as fast-growing and wood properties of *Eucalyptus cloeziana*.

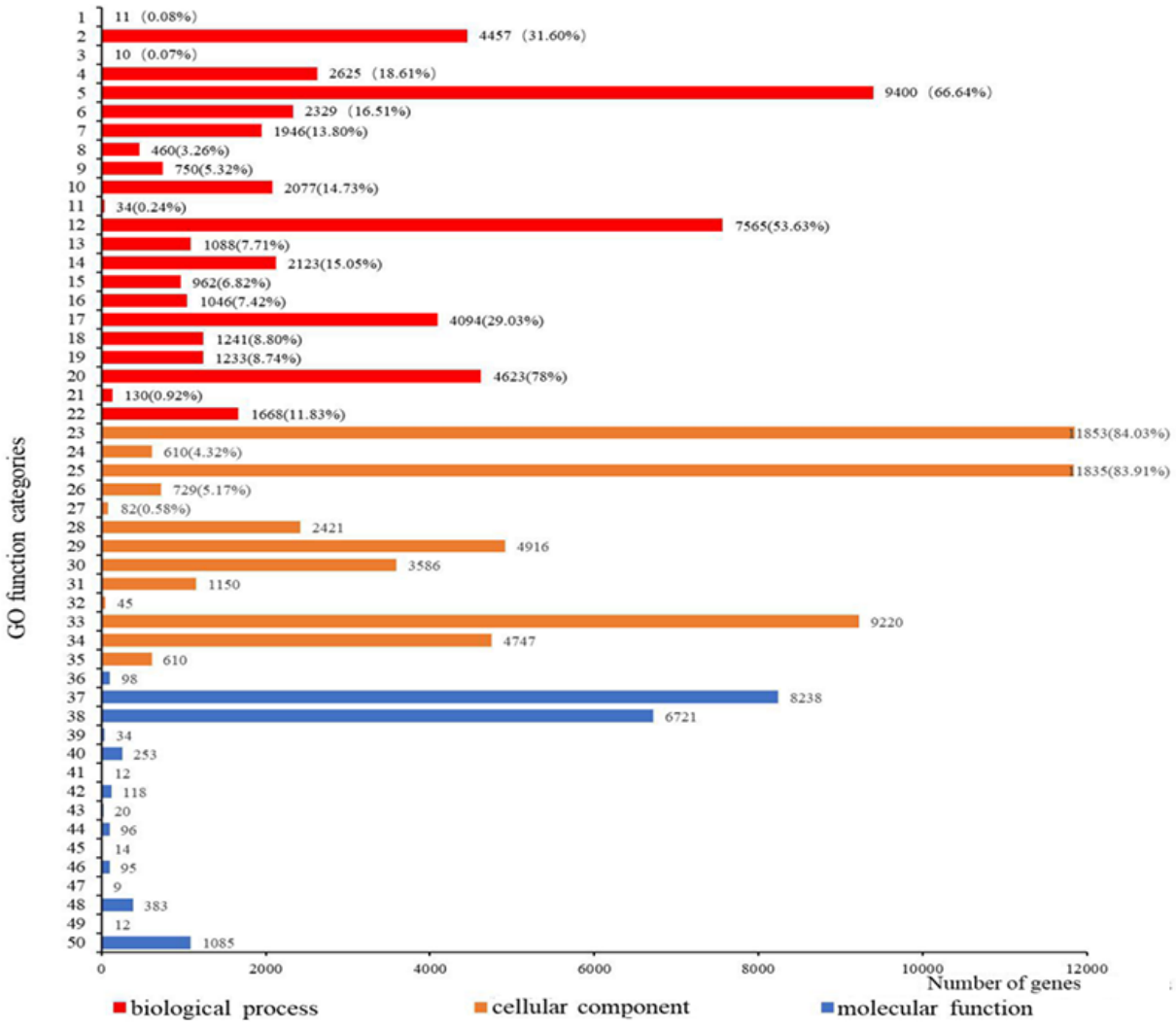

Figure 5 Gene Ontology of *Eucalyptus cloeziana* Unigene

Note: 1: Biological adhesion; 2: Biological regulation; 3: Cell killing; 4: Cellular component organization or biogenesis; 5: Cellular process; 6: Developmental process; 7: Establishment of localization; 8: Growth; 9: Immune system process; 10: Localization; 11: Locomotion; 12: Metabolic process; 13: Multi-organism process; 14: Multicellular organismal process; 15: Negative regulation of biological process; 16: Positive regulation of biological process; 17: Regulation of biological process; 18: Reproduction; 19: Reproductive process; 20: Response to stimulus; 21: Rhythmic process; 22: Signaling; 23: Cell; 24: Cell junction; 25: Cell part; 26: Extracellular region; 27: Extracellular region part; 28: Macromolecular complex; 29: Membrane; 30: Membrane part; 31: Membrane-enclosed lumen; 32: Nucleoid; 33: Organelle; 34: Organelle part; 35: Symplast; 36: Antioxidant activity; 37: Binding; 38: Catalytic activity; 39: Electron carrier activity; 40: Enzyme regulator activity; 41: Metallochaperone activity; 42: Molecular transducer activity; 43: Nutrient reservoir activity; 44: Protein binding transcription factor activity; 45: Protein tag; 46: Receptor activity; 47: Receptor regulator activity; 48: Structural molecule activity; 49: Translation regulator activity; 50: Transporter activity

1.2.5 Gene transcription factors (TFs)

Transcription factors (TFs) are proteins that bind to specific DNA sequences, which play a key role in the regulation of gene expression by controlling the transcription of genetic information from DNA to RNA. Among 22 099 Unigenes with annotation information, 1 021 were annotated to the transcription factor database, distributed in 65 families (Figure 7). In the family, bHLH (68), MYB (65), MYB-related (62), NAC (59) and C2H2 (53) accounted for the largest proportion of 6.66%, 6.37%, 6.07%, 5.78% and 5.19% respectively (Figure 7). The transcription factor family such as AP2/ERF-RAV, HRT, S1Fa-like, STAT and ULT had only one Unigene. TFs plays an important role in cell activities, especially bHLH (Yu et al., 2019), MYB (Chen et al., 2019) and other transcription factors play an important role in cell signal transduction and plant stress resistance.

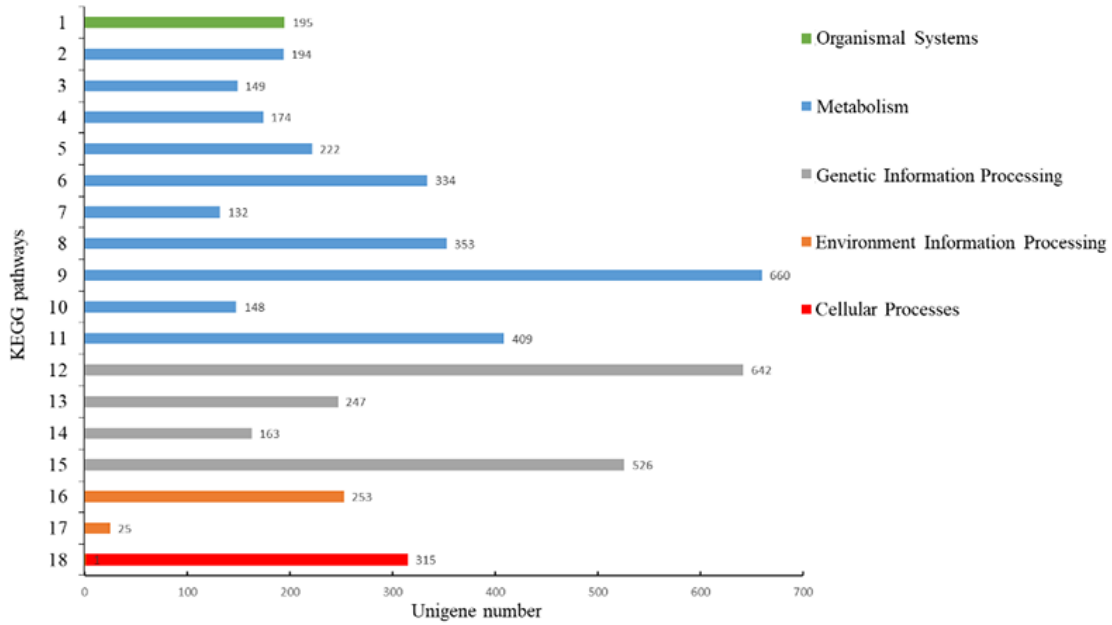

Figure 6 KEGG pathways of *Eucalyptus cloeziana* Unigene

Note: 1: Environmental adaptation; 2: Nucleotide metabolism; 3: Metabolism of terpenoids and polyketides; 4: Metabolism of other amino acids; 5: Metabolism of cofactors and vitamins; 6: Lipid metabolism; 7: Glycan biosynthesis and metabolism; 8: Energy metabolism; 9: Carbohydrate metabolism; 10: Biosynthesis of other secondary metabolites; 11: Amino acid metabolism; 12: Translation; 13: Transcription; 14: Replication and repair; 15: Folding, sorting and degradation; 16: Signal transduction; 17: Membrane transport; 18: Transport and catabolism

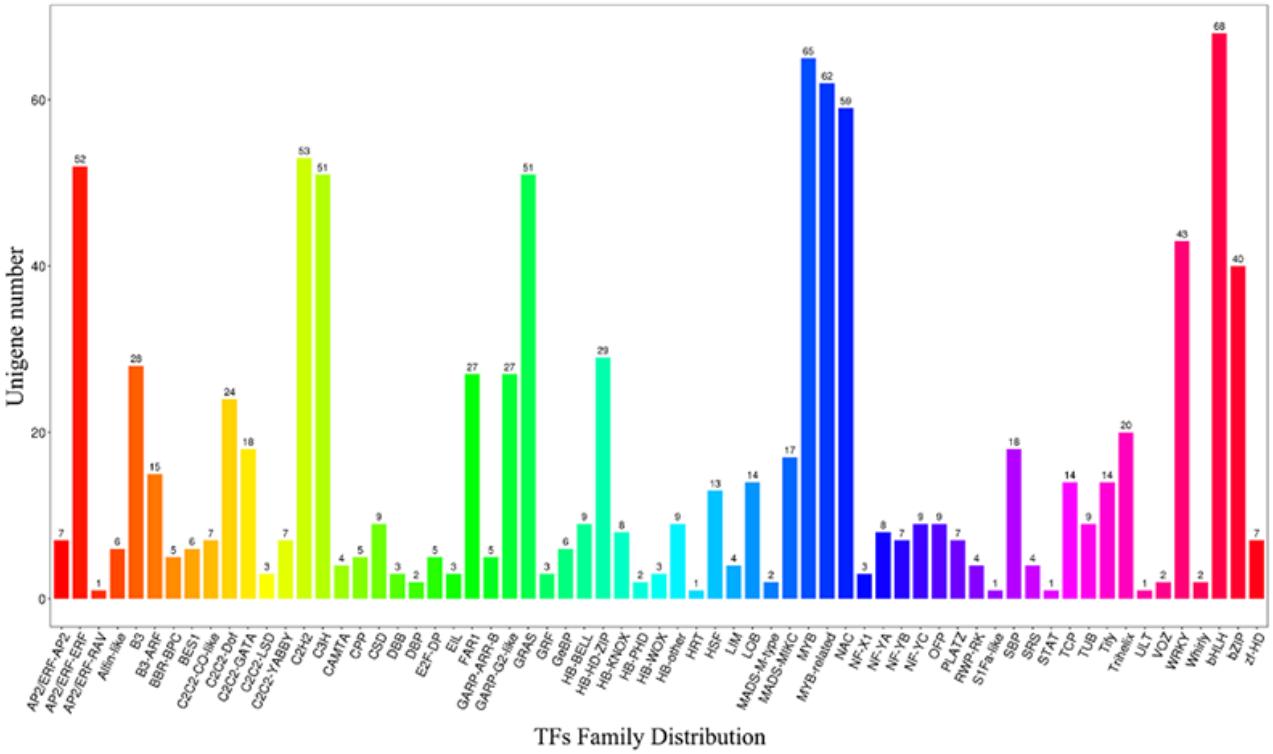

Figure 7 Transcription factor family distribution of *Eucalyptus cloeziana* Unigene

1.2.6 Plant resistance gene (PRG)

Plant resistance gene (PRG) (http://prgdb.crg.eu /wiki/Main_Page) database contains more than 112 disease resistance genes and 104 335 candidate disease resistance genes. This time, a total of 3 274 UniGenes annotated to the PRG database, distributed in 13 classes. The results of PRG comparison showed that the largest number of matched genes were RLP (785) and TNL (712), accounting for 23.98% and 21.74% of the total annotations, respectively; Followed by CNL (539) and NL (527), accounting for 16.46% and 16.10% of the total annotations, respectively (Figure 8). In the normal growth environment and state, RLP and TNL resistance gene families were specifically amplified in *Eucalyptus cloeziana* terminal buds, which may indicate that *Eucalyptus cloeziana* had potential strong adaptability and stress resistance, which provided a reference for the subsequent research on the stress resistance of *Eucalyptus cloeziana* to environmental and biological stresses, and also provided enlightenment for the resistance breeding of *Eucalyptus cloeziana* in the future.

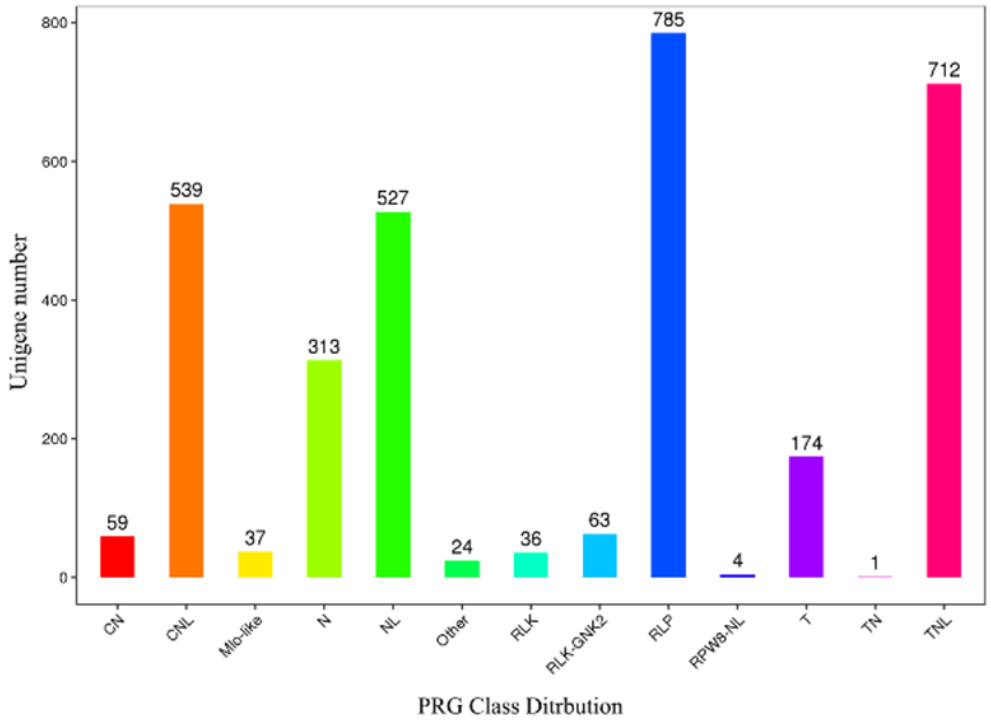

Figure 8 The PRG distribution of *Eucalyptus cloeziana* Unigene

1.3 Distribution characteristics of SSR loci in *Eucalyptus cloeziana* terminal buds

A total of 12 366 SSR loci were found from 26 587 Unigenes, which were distributed in 8 218 UniGene sequences, and the frequency was 46.51%, with an average of one SSR locus per 2.75 kb. 116 repeat motifs were found in the transcriptome of *Eucalyptus cloeziana* terminal buds, of which dinucleotide and trinucleotide were the main repeat motifs, accounting for 46.6% and 34.20% of the total SSRs respectively; The dominant repeat motifs were AG/CT and CCG/CGG, accounting for 44.94% and 12.53% respectively; SSR repeat number was mainly low repeat number (5~10 times) (71.28%); The average length of SSR motifs was 19.7 bp, and 63.75% of the motif length was concentrated between 12 and 20 bp. On the whole, *Eucalyptus cloeziana* transcriptome had high frequency of SSR, rich types of repeat motifs, and high repeat number (10 or more) accounted for a large proportion (38.8%), which had great development potential (Table 1).

Table 1 Type and repeat motifs of SSR loci in *Eucalyptus cloeziana*

Repeats	Repeat number								Total
	5	6	7	8	9	10	11	>11	
1 bp	0	0	0	0	0	678	340	1047	2065
2 bp	0	1105	851	651	600	453	398	1705	5763
3 bp	1796	1015	663	349	225	120	28	33	4229
4 bp	96	43	17	6	1	1	0	0	164
5 bp	36	10	0	1	0	0	0	0	47
6 bp	79	12	7	0	0	0	0	0	98
Total	2007	2185	1538	1007	826	1252	766	2785	12366
Proportion (%)	16.23	17.67	12.44	8.14	6.68	10.12	6.19	22.52	100

Note: The repeats 1~6 bp means mono-, di-, tri-, tetra-, penta- and hexa-nucleotide respectively

2 Discussion

As a valuable timber tree species, *Eucalyptus cloeziana* has broad development prospects. Abundant terminal bud transcriptome information obtained by high-throughput sequencing can provide rich resources for molecular assisted breeding of *Eucalyptus cloeziana*. In this study, through high-throughput sequencing and analysis of cDNA library of *Eucalyptus cloeziana* terminal buds, a total of 56 157 538 Raw reads were obtained, and a total of 8.4 Gb sequencing data were obtained. The Q30 was more than 93.99%, the GC content was 49.97%, and the base number distribution was stable. After splicing and assembly, a total of 26 587 high-quality Unigenes were obtained, with a total length of 34 023 kb, an average length of 1 279.69 bp, the length range from 301 to 1000 bp (13 527, accounting for 50.87% of the total), and N50 of 1851 bp. Based on the same sequencing technology, 53 433 Unigenes were obtained from the transcriptome sequences of *Eucalyptus cloeziana* roots. The length of N50 was 1 587 bp, the average length was 890 bp, the content of G+C was 46.86%, and 50.73% of Unigenes were concentrated between 200 and 500 bp (Zhu et al., 2018); The transcriptome of the near source genus *Syzygium samarangense* (Wei et al., 2018) was assembled to obtain 87 538 Unigenes, with an average length of 858.56 bp, and 75.84% of the Unigenes distributed between 300 and 1 000 bp; Based on the same platform, the average length of transcriptome sequences of *Pseudolarix amabilis* leaves was 699 bp and the length of N50 was 1 300 bp (Zhang, 2019); The average length of transcriptome of *Corylus avellana* axillary buds was 1 189 bp, and the length of N50 was 1 916 bp (Kavas et al., 2019); The average length of transcriptome Unigene of *Camellia sinensis* mature seed was 854 bp and the length of N50 was 1 480 bp (Jin et al., 2018). In comparison, the sequence of *Eucalyptus cloeziana* terminal buds was longer, the length distribution was more uniform, and the integrity of sequence assembly was better.

In this experiment, all Unigenes (26 587) of *Eucalyptus cloeziana* terminal buds were compared with seven public databases, and 83.12% (22 099) sequences were successfully annotated. It was much higher than 64.94% of the transcriptome of *Eucalyptus cloeziana* roots (Zhu et al., 2018), which may be related to the short sequence fragment of the latter and the lack of annotation information of the root. Through the analysis of NR database, the matching similarity between *Eucalyptus cloeziana* and *Eucalyptus grandis* was very high (94.55%), which was consistent with the NR comparison results of roots, which may be related to the integrity of biological information data of *Eucalyptus grandis*; In the GO database, 14 105 Unigenes were annotated and matched to 50 functional gene regions in three classes: biological function, cell component and molecular function. The regions executing biological processes accounted for the largest proportion, with a total of 49 872 annotation information, of which the number involved in cellular process and metabolic process was the largest, indicating that cell division and differentiation and metabolic activities were vigorous in the process of growth and differentiation of *Eucalyptus cloeziana* terminal buds. A large number of studies showed that cellular process and metabolic process were the most significant subclasses of GO functional biological process, especially in buds (Bian et al., 2019; Chang et al., 2019; Kavas et al., 2019); 7 117 Unigenes were annotated to KEGG function and were involved in 127 metabolic pathways, of which the genes involved in carbohydrate metabolism (660, 16.03%) accounted for the largest proportion, indicating that the apical meristem region of *Eucalyptus cloeziana* carried out frequent mitosis and vigorous physiological and metabolic activities.

Transcription factors (TFs) are proteins that can bind to specific DNA sequences and play a key role in the regulation of gene expression. In this study, compared with TF database, a total of 1 021 Unigenes were annotated into the transcription factor database and distributed in 65 families, of which bHLH (68) accounted for the largest proportion, followed by MYB (65), MYB-related(62), NAC(59) and C2H2(53) families. bHLH is one of the largest transcription factor families in plants. Yu et al. (2019) reviewed in detail that bHLH transcription factors play an important role in biological processes such as plant signal transduction, growth and stress resistance. MYB protein is the largest and most diverse transcription factor family in plants, which is mainly involved in cell cycle regulation, tissue and organ morphogenesis, secondary metabolite synthesis and stress resistant biological processes (Chen et al., 2019); Araújo (2018) revealed that the expression of MYB and NAC family members in the stems of *Eucalyptus grandis* and *Eucalyptus globulus* was significantly higher than that of the control under low temperature stress. This study also found that a total of 3 274 Unigenes were annotated into 13 resistance gene categories in PRG database, of which RLP was the largest, accounting for 23.98% of the total annotation. Activation of RLP protein (RuBisCO-like protein) usually leads to rapid accumulation of active oxygen species (AOS), changes in cell ion flux, activation of protein kinase cascade, changes in gene expression, and possible targeted protein degradation (Tabita et al., 2008). A large amount of information obtained through transcriptome analysis will provide a strong basis for the study on growth and stress resistance of *Eucalyptus cloeziana*.

SSR molecular marker is an ideal marker to study the polymorphism of plant genetic resources, molecular marker assisted breeding and genetic mapping. In this study, 12 366 SSR loci were obtained from 26 587 Unigenes in the transcriptome of *Eucalyptus cloeziana* terminal buds, with a frequency of 46.51% and a distribution density of 1/2.75 kb. Dinucleotide and trinucleotide were the main repeat motifs, accounting for 46.6% and 34.20% of the total SSR respectively. The distribution density of SSR loci was consistent with the analysis results of 71 115 EST sequences based on database *Eucalyptus* (*E. grandis*, *E. globosus*, *E. saligna* 和 *E. urophylla*) reported by Ceresini et al. (2005): 20 530 SSR loci with an average distribution distance of 1/2.7 kb; The distribution density of SSR loci was higher than that of other *Eucalyptus* transcriptome (1/3.7 kb) reported by He et al. (2015), and the frequency of SSR in *E. globulus* transcriptome was 23.15% reported by Acuña et al. (2012); The distribution density of SSR loci was lower than that of *Camellia sinensis* (1/2.61 kb) (Jin et al., 2006) and *Hevea brasiliensis* (1/2.25 kb, the frequency of SSR was 63.71%) (Feng et al., 2009). In general, SSR loci in transcriptome of *Eucalyptus cloeziana* terminal buds have high frequency, high distribution density and rich types of repeat motifs, which will have great application potential in genetic diversity research, molecular marker assisted breeding and genetic mapping of *Eucalyptus cloeziana*.

3 Materials and Methods

3.1 Test materials

The test materials were the stem terminal buds of 2-year-old *Eucalyptus cloeziana*, which were collected from the *Eucalyptus cloeziana* Test Base of Dongmen Forest Farm, Fusui County, Chongzuo City, Guangxi Zhuang Autonomous Region (22°17'22.30"N, 107°14'108.00"E). After the sample was collected, it was quickly put into liquid nitrogen quick freezing for preservation. In October 2019, total RNA was extracted and cDNA library was constructed by the OEbiotech, and the original reads were obtained through Illumina HiSeq X Ten high-throughput sequencing platform.

3.2 Assembly and splicing of transcriptome sequence

The raw data obtained in the sequencing process were tested, filtered and quality controlled to obtain high-quality clean reads, and then De-novo splicing and assembly was carried out based on Trinity software (Grabherr et al., 2011), that is, under the condition of not relying on the reference genome, the reads with overlap were connected into a longer sequence and spliced into Contigs after continuous extension, according to the sequence similarity and length, the longest one was selected as Unigene, and then bioinformatics statistical analysis was carried out.

3.3 Function annotation of transcriptome

By using diamond (Buchfink et al., 2015) software, all Unigene sequences of *Eucalyptus cloeziana* were annotated and compared with NR (nonredundant protein database, https://www.ncbi.nlm.nih.gov/), KOG (karyotic orthologous groups, http://www.ncbi.nlm.nih.gov/COG), GO (gene ontology, http://www.geneontology.org), Swiss-Prot (Swiss Prot protein database, http://web.expasy.org/docs/swiss-prot_guideline.html), eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups, http://eggnog.embl.de/version_3.0/), KEGG (Kyoto encyclopedia, http://www.genome.jp/kegg/) biological information database; The predicted amino acid sequence of Unigene was aligned to Pfam database (http://pfam.xfam.org/) by using HMMER (E-value ≤ 1e-10) software (Mistry et al., 2013) to obtain the function annotation information of Unigene; By Using iTAK (Zheng et al., 2016), Unigene was aligned to its own database to obtain the annotation information of transcription factors (TFs); The Unigene sequence was aligned to PRG (plant resistance gene) database, and the best one with E value less than 1e-5 was screened as the annotation information of the Unigene.

3.4 SSR analysis of transcriptome

SSR search and statistical analysis were performed on all Unigenes by using MISA (MIcroSAtellite identification tool) (http://pgrc.ipk-gatersleben.de/misa/). The SSR search conditions were set as follows: the repeat motif was 1~6 bp, meaning mono-, di-, tri-, tetra-, penta- and hexa-nucleotide respectively. The minimum repeat number was 10, 6, 5, 5, 5 and 5 times respectively. The recognition condition of composite SSR was that the distance between two SSR sites was ≤100 bp.

Authors' Contributions

LJ and JWX jointly completed the experimental design, data analysis and the writing of the first draft of the manuscript; ZL and LXY participated in sampling and partial data analysis; BTD was the conceiver and person in charge of the project, guiding the experimental design, data analysis, manuscript writing and revision. All authors read and approved the final manuscript.

Acknowledgments

This study was jointly funded by Key Research and Development Project of Guangxi (2018AB44025) and the Scientific Research Project of Guangxi University "Research on Integrated Technology of *Eucalyptus cloeziana* Plantation" (20180493).

References

Acuña C.V., Fernandez P., Villalba P.V., García M.N., Hopp H.E., and Poltri S.N.M., 2012, Discovery, validation, and in silico functional characterization of EST-SSR markers in *Eucalyptus globulus*, Tree Genetics & Genomes, 8: 289-301
https://doi.org/10.1007/s11295-011-0440-0

Araújo P., 2018, Stem transcriptome of cold stressed *Eucalyptus globulus* and *E. urograndis*, Trends in Horticulture, 1(2): 1e
https://doi.org/10.24294/th.v1i2.889

Bian X., Qu C., Zhang M., Li Y., Han R., Jiang J., and Liu G., 2019, Transcriptome sequencing to reveal the genetic regulation of leaf margin variation at early stage in birch, Tree Genetics & Genomes, 15: 4
https://doi.org/10.1007/s11295-018-1312-7

Buchfink B., and Xie C., and Huson D.H., 2015, Fast and sensitive protein alignment using DIAMOND, Nature Methods, 12(1): 59-60
https://doi.org/10.1038/nmeth.3176
PMid:25402007

Ceresini P.C., Silva C.L., Missio R.F., Souza E.C., Fischer C.N., Guillherme I.R., Gregorio L., da Silva E.H.T., Cicarelli R.M.B., da Silva M.T.A., Garcia J.F., Avelar G.A., Neto L.R.P., Marçon A.R., Junior M.B., and Marini D.C., 2005, Satellyptus: analysis and database of microsatellites from ESTs of *Eucalyptus*, Genetics and Molecular Biology, 28(3): 589-600
https://doi.org/10.1590/S1415-47572005000400014

Chang Y., Hu T., Zhang W., Zhou L., Wang Y., and Jiang Z., 2019, Transcriptome profiling for floral development in reblooming cultivar 'High Noon' of *Paeonia suffruticosa*. Scientific Data 6: 217
https://doi.org/10.1038/s41597-019-0240-1
PMid:31641161 PMCid:PMC6805890

Chen C., Zhang K., Khurshid M., Li J. and Zhou M., 2019, MYB transcription repressors regulate plant secondary metabolism, Critical Reviews in Plant Sciences, 38(3): 159-170
https://doi.org/10.1080/07352689.2019.1632542

Deng Z.Y., Chen J.B., Guo D.Q, Li C.R., and Lu C.X., 2019, Genetic diversity of *Eucalyptus cloeziana*, Linye Kexue Yanjiu (Forest Research), 32(4): 41-46

Feng S.P., Li W.G., and Huang H.S., 2009, Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*), Molecular Breeding, 23(1): 85-97
https://doi.org/10.1007/s11032-008-9216-0

Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., and Regev A., 2011, Full-length transcriptome assembly from RNA-Seq data without a reference genome, Nat. Biotechnol., 29(7): 644-652
https://doi.org/10.1038/nbt.1883
PMid:21572440 PMCid:PMC3571712

He X., Zheng J., Zhou J., He K., Shi S., and Wang B., 2014, Characterization and comparison of EST-SSRs in *Salix*, *Populus*, and *Eucalyptus*, Tree Genetics & Genomes, 11(820): 1-10
https://doi.org/10.1007/s11295-014-0820-3

Huang Z., Zhang J., Chen Z., Wang L.H., and Guo H.Y., 2018, Development and prospects of heredity and breeding researches on *Eucalyptus cloeziana*, Sichuan Linye Keji (Journal of Sichuan Forestry Science & Technology), 39(1): 17-21

Jin J.Q., Cui H.R., Chen W.Y, Lu M.Z., Yao Y.L., Xin Y., and Gong X.C., 2006, Data mining for SSRs in ESTs and development of EST-SSR marker in tea plant (*Camellia sinensis*), Chaye Kexue (Journal of Tea Science), 26(1): 17-23

Jin X., Liu D., Ma L., Gong Z., Cao D., Liu Y., Li Y., Jiang C., 2018, Transcriptome and expression profiling analysis of recalcitrant tea (*Camellia sinensis* L.) seeds sensitive to dehydration, International Journal of Genomics, https://doi.org/10.1155/2018/ 5963797
https://doi.org/10.1155/2018/5963797
PMid:29967765 PMCid:PMC6008840

Kavas M., Kurt Kızıldoğan A., and Balık H., 2019, Gene expression analysis of bud burst process in European hazelnut (*Corylus avellana* L.) using RNA-Seq, Physiol Mol Biol Plants, 25(1): 13-29
https://doi.org/10.1007/s12298-018-0588-2
PMid:30804627 PMCid:PMC6352538

Li C.R., Chen K. and Zhou X.J., 2012a, Research status and development trend of *Eucalyptus cloeziana*, Anshu Keji (Eucalypt Science & Technology), 29(2): 40-46

Li C.R., Xiang D.Y., Chen J.B., Zhai X.C., Kan R.F., and Lan J., 2012b, Study on basic density variation of *Eucalyptus cloziana*, Zhongnan Linye Keji Daxue Xuebao (Journal of Central South University Forestry & Technology), 32(6): 158-162

Lü J., Li C., Zhou C., Chen J., Li F., Weng Q., Li M., Wang Y., Chen S., Chen J., and Gan S., 2020, Genetic diversity analysis of a breeding population of *Eucalyptus cloeziana* F. Muell. (Myrtaceae) and extraction of a core germplasm collection using microsatellite markers, Industrial Crops and Products, 13-29
https://doi.org/10.1007/s12298-018-0588-2
PMid:30804627 PMCid:PMC6352538

Mistry J., Finn R.D., Eddy S.R., Bateman A., and Punta M., 2013, Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions, Nucleic Acids Research, 41(12): e121
https://doi.org/10.1093/nar/gkt263
PMid:23598997 PMCid:PMC3695513

Steane D.A., Nicolle D., and Sansaloni C.P., 2011, Population genetic analysis and phylogeny reconstruction in *Eucalyptus* (Myrtaceae) using high-throughput, genome-wide genotyping, Molecular Phylogenetics & Evolution, 59(1): 206-224
https://doi.org/10.1016/j.ympev.2011.02.003
PMid:21310251

Tabita F.R., Hanson T.E., Li H., Satagopan S., and Chan S., 2008, Function, structure, and evolution of the RubisCO-Like proteins and their RubisCO homologs, Microbiology & Molecular Biology Reviews, 71(4): 576-599
https://doi.org/10.1128/MMBR.00015-07
PMid:18063718 PMCid:PMC2168653

Vaattovaara A., Leppala J., Salojarvi J., and Wrzaczek M., 2019, High-throughput sequencing data and the impact of plant gene annotation quality, Journal of Experimental Botany, 70(4): 1069-1076
https://doi.org/10.1093/jxb/ery434
PMid:30590678 PMCid:PMC6382340

Wang J.Z., Xiong T., Zhang L., Li Q.W., Fei X.Y., Shi Q., Li H.L., and Lan J., 2016, Genetic variation and selection on growth and stem form quality traits of 25-year-old *Eucalyptus cloeziana* provenance, Linye Kexue Yanjiu (Forest Research), 29(5): 705-713

Wei X.Q., Xu L., Zhang X.J., Yu D., Chen Z.F., Zhang L.M., Chen C.Z., and Xu J.H., 2018, Analysis on SSR information in transcriptome and development of molecular markers in wax apple, Yuanyi Xuebao (Acta Horticulture Sinica), 45(3): 541-551

Xiang D.Y., Chen J.B., Shen W.H., Zhou W. and Kan R.F., 2008, Research on wood physical property variation among provenances of *Eucalyptus cloeziana*, Guangxi Linye Kexue (Guangxi Forestry Science), 37(2): 3-11

Yu B., Tian Y., Li H.Y., Lv X.Y., Wang Y.G., and Duanmu H.Z., 2019, Research progress of plant bHLH transcription factor, Zhongguo Nongxue Tongbao (Chinese Agricultural Science Bulletin), 35(9): 75-80

Zhang W., Zhang L., Kou Y., and Zhang Z., 2019, Transcriptome sequencing and bioinformatic analysis of *Pseudolarix amabilis*, an endangered gymnosperm in China, Acta Agriculturae Universitatis Jiangxiensis, 41(4): 761-772

Zheng Y., Jiao C., Sun H., Rosli H.G., Pombo M.A., Zhang P., and Banf M., 2016, iTAK: a program for genome-wide prediction and classification of plant transcription factors,transcriptional regulators, and protein kinases, Molecular Plant, 9(12): 1667-1670
https://doi.org/10.1016/j.molp.2016.09.014
PMid:27717919

Zhu L., Guo L., He S., Hui L., Liu X., and Chen S., 2018, Transcriptome characterization analysis of *Eucalyptus cloeziana* root based on Illumina HiSeq 2000 sequencing technology, Molecular Plant Breeding, 16(13): 4245-4254