



Original Article

Mining lycodine-type alkaloid biosynthetic genes and genetic markers from transcriptome of *Lycopodium casuarinoides*Gang Chen^{a,b}, Qiu-mei Lin^a, Lin Zeng^a, Yi-ping Zou^{a,*}^aKey Laboratory of Natural Active Pharmaceutical Constituents, College of Chemistry and Biology Engineering, Yichun University, Yichun 336000, China^bCollege of Life Sciences and Resource Environment, Yichun University, Yichun 336000, China

ARTICLE INFO

Article history:

Received 9 July 2019

Revised 2 September 2019

Accepted 10 September 2019

Available online 27 March 2020

Keywords:

Lycopodium casuarinoides (spring) Holub ex Dixitlycodine-type alkaloids biosynthesis
simple sequence repeat (SSR)
transcriptome

ABSTRACT

Objective: *Lycopodium casuarinoides*, a fern of the Lycopodiaceae family, is a traditional Chinese medicine, which has similar efficacy to that of *Huperzia serrata* in treating rheumatoid arthritis (RA). However, they are different in the contents and compositions of lycopodium alkaloids. In this study, the biosynthesis related genes of lycopodium alkaloids and genetic markers are discovered in *L. casuarinoides* transcriptome.

Methods: The plant of *L. casuarinoides* was collected and was subjected to the RNA isolation, cDNA library construction, high throughput RNA sequencing and bioinformatics analysis.

Results: Totally 124,524 high-quality unigenes were assembled from RNA sequencing reads, with an average sequence length of 601 bp. Among the *L. casuarinoides* transcripts, 61,304 shared the significant similarity (E-value < 10⁻⁵) with existing protein sequences in the public databases. From 124,524 unigenes, 47,538 open reading frames (ORFs) were predicted. Based on the bioinformatics analysis, all possible enzyme genes involved in the lycodine-type alkaloids biosynthetic pathway of *L. casuarinoides* were identified, including lysine decarboxylase (LDC), primary amine oxidase (PAO), malonyl-CoA decarboxylase, etc. Sixty-four putative cytochrome p450 (CYP) and 827 putative transcription factors were selected from the transcriptome unigenes as the candidates of lycodine-type alkaloids biosynthesis modifiers. Furthermore, 13,352 simple sequence repeats (SSRs) were identified from 124,524 unigenes, of which dinucleotide motifs AG/CT were the most abundant (50.1%). Meanwhile, we confirmed the amplification effectiveness of 25 PCR primer pairs for randomly selected SSRs.

Conclusion: We obtained the comprehensive transcriptomic information from the high throughput RNA sequencing and bioinformatics analysis, which provided a valuable resource of transcript sequences of *L. casuarinoides* in public databases.

© 2020 Tianjin Press of Chinese Herbal Medicines. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license.
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

1. Introduction

Lycopodium casuarinoides (Spring) Holub ex Dixit is a perennial fern belonging to the Lycopodiaceae family (sensu lato), and is widely distributed in southern China. The whole plant of *L. casuarinoides*, also called as Shujincao in Chinese, is commonly used as the folk medicine against rheumatoid arthritis (RA). Recent pharmacological studies have shown that the total alkaloids of *L. casuarinoides* (ALC) are a potential agent for treating the inflammation and arthritis. Evidence from laboratory and clinical trials

showed that the total alkaloids of *L. casuarinoides* possess multiple pharmacological activities, including anti-inflammation and anti-arthritis (Pan, Xia, Guo, & Kong, 2015), anti-neuronal cell damage (Tang et al., 2013), and anti-acetylcholinesterase (AChE) (Hirasawa et al., 2008; Tang et al., 2013; Zhang, Chen, Song, Zhang, & Gao, 2014).

The chemical structures, composition properties of lycodine-type alkaloids and their derivatives of *L. casuarinoides* have been studied extensively and are well defined (Hirasawa et al., 2008; Ma & Gang, 2004; Tang et al., 2013). There are four types of lycopodium alkaloids, i.e., lycopodine-type, miscellaneous-type, fawcettimine-type, and lycodine-type, and lycodine-type compounds have been identified in *L. casuarinoides* (Ma & Gang, 2004).

* Corresponding author.

E-mail address: zyping66@163.com (Y.-p. Zou).

The lycodine-type alkaloids represent a unique class of compounds that are characterized by three or four connected hexatomic rings (Ma, Jiang, & Zhu, 1998). The hexatomic ring is composed of four parts including a pyridine ring (called ring A), a piperidine ring (ring B) and a bicyclononane core formed by rings B and D (Ma & Gang, 2004; Tang et al., 2013; Zhang et al., 2014).

To date, more than 80 lycodine-type alkaloids along with their known analogues have been isolated from *L. casuarinoides*, such as huperzine B, huperzine C, and lycoparin C, all of which have been found to possess the potent inhibition ability against AchE. Interestingly, it is proved that AchE inhibiting activity of huperzine C is comparable to that of huperzine A (HupA), while the latter has been clinically used to treat Alzheimer's disease (AD) for many years in China (Hirasawa et al., 2008; Tang et al., 2013; Zhang et al., 2014).

Ma and Gang (2004) proposed Hup A and *lycopodium* alkaloids biosynthetic pathways for the first time. The biosynthetic pathways of lycodine-type alkaloids were also well documented in this proposed pathway. In the studies of two transcriptomes of *Huperiza serrata* (Thunb. ex Murray) Trev. and its closely related species, *Phlegmariurus carinatus* (Desv.) Ching (Luo et al., 2010a; Yang et al., 2017), the biosynthetic pathways of *lycopodium* alkaloids are well summarized as follows: initially, the first biosynthetic step is l-lysine converting to cadaverine, catalyzed by lysine decarboxylase (LDC). Subsequently, the oxidative deamination of cadaverine is catalyzed by copper amine oxidase (CAO), and then piperidine is spontaneously generated. Next, piperidine is reacted with oxoglutaric acid (produced from malonyl-CoA), which is followed by decarboxylation, and then pelletierine is generated. After that, multi-step reactions result in the biosynthesis of lycodane that contains the unique tetracyclic skeleton. Finally, four classes of *lycopodium* alkaloids, including lycopodine-type, lycodine-type, fawcettimine-type and miscellaneous-type compounds, can be formed from lycodane via divergent chemical processes, such as the modification and oxidation, which are probably catalyzed by CYPs (Yang et al., 2017).

H. serrata, *P. carinatus* and *L. casuarinoides* are rich in *lycopodium* alkaloids (Hirasawa et al., 2008; Jiang et al., 2014; Ma et al., 1998), but they are distinguished in the contents and compositions of these alkaloids. Therefore, it is worth exploring whether *L. casuarinoides* has similar genes related to the biosynthesis of *lycopodium* alkaloids as above-mentioned.

Next-generation sequencing technologies (NGS) have been widely applied in the *de novo* transcriptome sequencing of medicinal plants without available genomic data (Sharma & Shrivastava,

2016). In this study, we carried out the high throughput transcriptome sequencing of the aerial tissues (stems and pin-like leaves) of *L. casuarinoides* and identified candidate genes involved in the lycodine-type alkaloids biosynthetic pathway. Totally 109 million sequencing reads, representing 16.43 Gb of clean data were utilized to assemble 124,524 unigenes. In addition, SSR-motifs (microsatellite markers) were also identified in the transcriptome dataset of *L. casuarinoides*. The current results provide a valuable genetic basis for the synthetic biology research on bioactive *lycopodium* alkaloids of traditional medicinal plants.

2. Materials and methods

2.1. Plant materials

The aerial part of *L. casuarinoides* plant was collected from Minyue Mountain area (27°36' N, 114°15' E), Wentang Town, Yichun City, Jiangxi Province of China in May 2016 (Fig. 1). A voucher specimen (No. 20160613) was deposited at the key Laboratory of Natural Active Pharmaceutical of Jiangxi Province, Yichun University, China. The sample was authenticated by Assoc. Prof. Chang-jiu Ji (College of Bioscience, Yichun University). The wild *L. casuarinoides* was grown in a wild and moist environment. In order to facilitate sampling and protect local germplasm resources, we used the aboveground part (stems and leaves) in the following analysis. All fresh samples were rapidly mixed, flash frozen in liquid nitrogen and stored at -80 °C until further use.

2.2. RNA preparation

The total RNA was extracted from *L. casuarinoides* samples by Plant RNA Isolation Kit (Tiangen Co., Ltd., Beijing, China); Phenolic compounds were eliminated and the genomic DNA was removed according to the manufacturer's directions. The RNA degradation and contamination were assessed on 1% agarose gel. The RNA quality was checked using the NanoPhotometer spectrophotometer (Implen, CA, USA). The RNA concentration was quantified using Qubit RNA Assay Kit in Qubit 2.0 fluorometer (Life Technologies, CA, USA), and the total RNA integrity (RIN value) was checked using RNA 6000 Nano Assay Kit with the Agilent bioanalyzer 2100 system (Agilent Technologies, CA, USA). The RIN of the above-mentioned sample was required to reach more than 7.5 for RNA sequencing.



Fig. 1. Medicinal plant *L. casuarinoides*. (A) Above-ground part of *L. casuarinoides*. (B) Stems and leaves at four different developmental stage (From left to right: adult stage, maturation stage, juvenile stage, initial stage, according to branching phenotype).

2.3. cDNA library construction and transcriptome sequencing

Equal amounts of total RNA from the aerial tissues, including main stems and pin-like leaves, were pooled and then were used to construct the cDNA library. The library was sequenced on the Illumina HiSeq™ 2500 platform in Novogene Co., Ltd. (Beijing, China) and the 100 bp paired-end RNA-Seq reads were generated based on the standard protocol (Illumina Inc., San Diego, CA). Raw reads of the sequencing data are deposited in NCBI Sequence Read Archive (SRA) with the accession number SRA2776088.

2.4. De novo assembly of transcriptome dataset and functional annotation of unigenes

The original image data generated by the sequencing platform were converted by base calling into raw reads. The raw reads were preprocessed by trimming adaptor sequences and discarding low quality reads that contain more than 50% bases with Q value of no more than 20. The clean reads were subjected to the *de novo* assembly of transcriptome using Trinity program (Grabherr et al., 2011). The contigs were obtained and were then further processed with the sequence clustering software, TGICL (Pertea et al., 2003) to generate longer sequences defined as unigenes. The BLAST software (a threshold *E*-value cutoff of 10^{-5}) was used to compare the assembled unigenes with those available from the public databases including NCBI Non-redundant (NR), Eukaryotic orthologous groups of proteins (KOG), Swiss-Prot, Protein family (Pfam), Gene ontology (GO), and the Kyoto Encyclopedia of Genes and Genomes (KEGG).

The functional annotation of unigenes was performed using the Blast2GO program (Conesa et al., 2005). In order to understand the distribution of gene functions, the WEGO (Ye et al., 2006) software was used to perform the GO analysis. In addition, the metabolic pathway mapping of transcripts was performed by KEGG (Kanehisa et al., 2008). The enzyme commission (EC) was assigned to the unique putative unigenes using the BLAST searching against the KEGG database.

The protein coding sequences (CDSs) of all unigenes were identified using BLASTX and ESTScan program (Iseli, Jongeneel, & Bucher, 1999). Briefly, the BLASTX alignment was performed between unigenes and protein databases such as NR, SwissProt, KEGG and KOG. The best alignment results were used to determine the sequences direction of unigenes. Unigenes with the sequence match in only one database were not further searched. When a

unigene did not align to any database, ESTScan was used to predict coding regions and determine the sequence direction.

Candidate genes belonging to the lycodine type alkaloids biosynthetic pathway were identified from the unigenes by database searching and functional annotation.

2.5. SSR analysis

In order to find useful SSRs from all unigenes of *L. casuarinoides*, the microsatellite identification tool MISA (<http://pgrc.ipk-gatersleben.de/misa/>) was used to search SSRs with the following parameters: the total repeat length of the sequence is longer than 20 bases, and the minimum repeats of di-, tri-, tetra-, penta- and hexa-nucleotides are 6, 5, 4, 4, and 4, respectively. A maximum distance of 100 nucleotides was allowed between two SSRs. The primers used for PCR verification of selected SSR loci were designed using Primer3.0 software (<http://primer3.ut.ee/>) and were listed in Datasheet 3. The expected PCR product size ranged from 100 to 280 bp, with the GC content ranging from 40% to 60%.

3. Results

3.1. Transcriptome sequencing output and de novo assembly of unigenes

To obtain the global information of *L. casuarinoides* transcriptome, the cDNA library was constructed from an equal mixture of total RNA isolated from main stems and pin-like leaves, and the high throughput transcriptome sequencing was performed. After quality check and data cleaning (Fig. S1), 109,546,778 high-quality clean reads were obtained, constituting a total length of 16.43 Gb of nucleotides. Among these clean reads, 97.49% of reads had Q20 bases (base quality more than 20), and the GC content was 45.78%. Because the *L. casuarinoides* genome information is still not available, all the clean reads were subjected to the *de novo* assembly using the Trinity program, from which 124,524 unigenes were obtained, with an average length of 601 bp, and 17,836 unigenes (14.32%) were longer than 1000 bp (Fig. 2). Among unigenes identified, 32.3% contained the CDS. Eighty percent (38,030) of CDSs were within the range of 201 to 1000 bp and the percentage of CDSs longer than 1000 bp was 20% (9508) (Fig. 2).

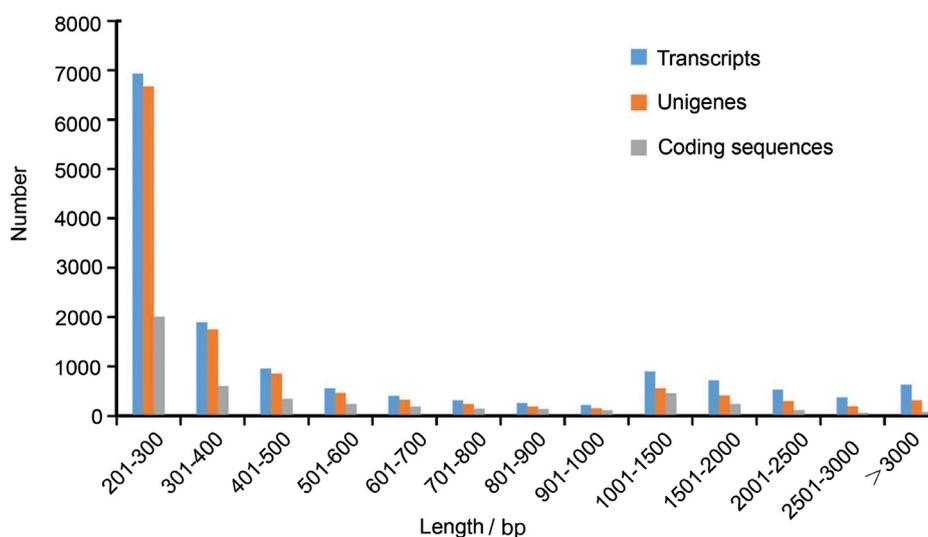


Fig. 2. Overview of *L. casuarinoides* transcriptome assembly and length distribution of CDS, unigene and transcript.

3.2. Unigene sequence annotation

To obtain the annotations of the assembled unigenes, we conducted a sequence similarity search against the seven public databases including NCBI Nr, NCBI Nt, SwissProt, KEGG, PFAM, GO, and KOG using BLASTX. Totally 61,304 unigenes (49.23%) were annotated against the public databases (Data Sheet 1, Supplementary materials), and 6783 unigenes were assigned with at least one public database; unigenes that were annotated are as follows: 44,816 unigenes in the Nr database, 43,590 in the SwissProt database, 26,217 in the KOG database, and 21,343 in the KEGG database (Table 1 and Fig. S2).

The results showed that 3.47% of unigenes shorter than 1000 bp had BLAST matches against the Nr database, whereas only 1.91% of unigenes longer than 1000 bp had BLAST matches (Fig. S3). The same tendency was observed in BLAST results against the Swiss-Prot database. The statistical analysis of the *E*-value features in the Nr database revealed that 23.6% of the mapped unigenes showed significant homology (E -value $<10^{-5}$) and 47.1% showed high similarity (60%–80%) to the available plant sequences. In the Nr database searching, the transcriptome sequences of *L. casuarinoides* showed high similarity to *Selaginella moellendorffii* (15.6%), *P. patens* (11.1%), *B. napus* (5.7%), and other species (4.0%) (Figs. S4 and S5). This indicates that the genome of *L. casuarinoides* is more

closely related to that of *S. moellendorffii* than to genomes of other four species, as both of them are ferns.

3.3. Functional classification by GO and KOG

For the functional annotation of the *L. casuarinoides* transcriptome, 43,080 unigenes were characterized using the GO analysis. In the biological process category, cellular processes (23,720, 55.06%), metabolic process (23,542, 22.27%), and single-organism process (19,069, 18.05%) were prominently represented. In the cellular component category, unigene sequences related to cell (13,703, 31.81%), cell part (13,642, 31.67%), and organelle (9313, 13.92%) were dominant. In the molecular function category, binding (21,231, 43.22%) and catalytic activity (19,686, 40.08%) were dominant (Fig. 3).

All unigenes were also subjected to the search against the KOG database for functional prediction and classification. In total, 50,534 unigenes were assigned to one or more of the 25 KOG classification categories. Among the 25 KOG categories, the largest cluster was posttranslational modification, protein turnover, chaperones (7297, 14.44%), followed by general function prediction (4726, 9.35%), transcription (4661, 9.22%), unknown functions (4347, 8.60%), replication, recombination and repair (4009, 7.93%), cell cycle control, cell division, and chromosome partitioning (3537, 7.00%), signal transduction mechanisms (2955, 5.85%), and finally cell wall/membrane/envelope biogenesis (2878, 5.70%) (Fig. 4). Only a few unigenes were assigned to the cell motility and extracellular structures. Additionally, 1074 unigenes were assigned to secondary metabolites, biosynthesis, transport and catabolism, of which 108 unigenes are related to the synthesis and metabolism of alkaloids. These GO and KOG annotations provide comprehensive information on specific biological processes, molecular functions, and cellular structures of *L. casuarinoides* transcripts, and may lead to the identification of novel genes involved in secondary metabolite biosynthesis pathways (Table S1).

3.4. Functional classification by KEGG

To analyze the gene products during metabolic processes and determine their functions in cellular processes in *L. casuarinoides*,

Table 1
Annotation percentage of *L. casuarinoides* unigenes against seven public databases.

Databases	Numbers of unigenes	Annotation percentage/%
Nr	44,816	35.98
Nt	19,171	15.39
Swiss-Prot	43,590	35
PFAM	41,395	33.24
KOG	26,217	21.05
KEGG	21,343	17.13
GO	43,080	34.59
Total no. of unigenes	124,524	–
Annotated in at least one database	61,304	49.23
Annotated in all databases	6783	5.44

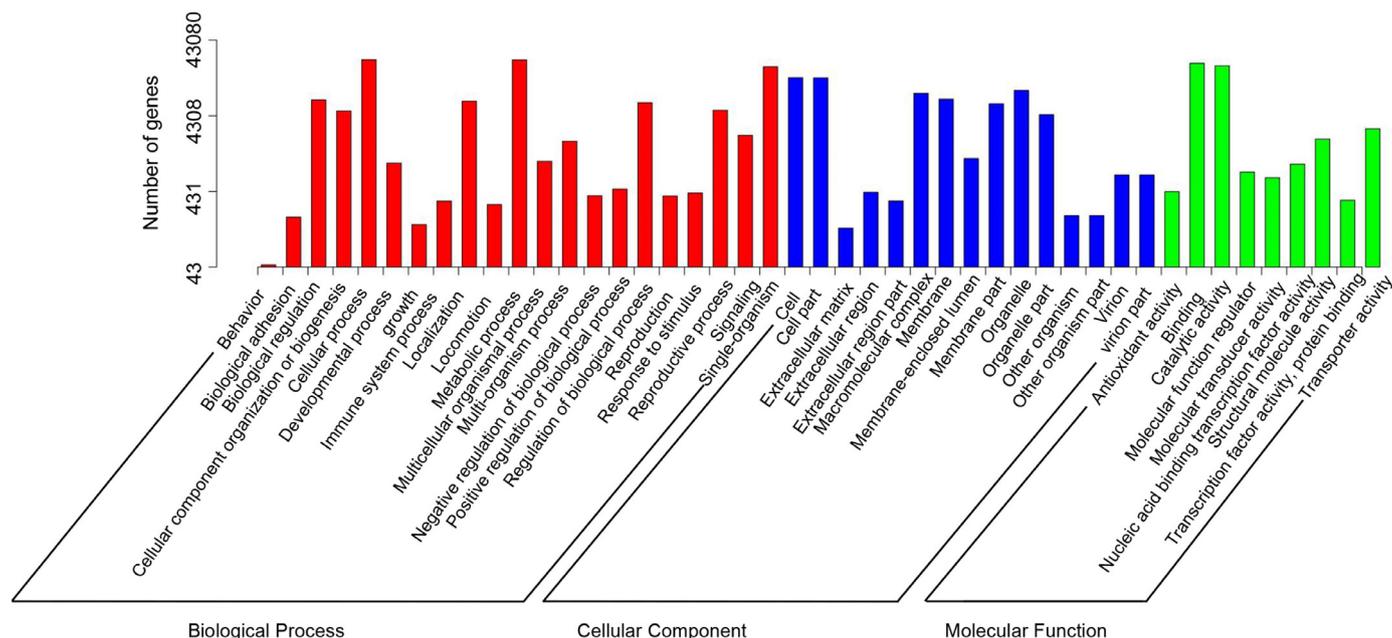


Fig. 3. GO classification of assembled unigenes. A total of 43,080 unigenes were categorized into three main categories: biological process, cellular component, and molecular function.

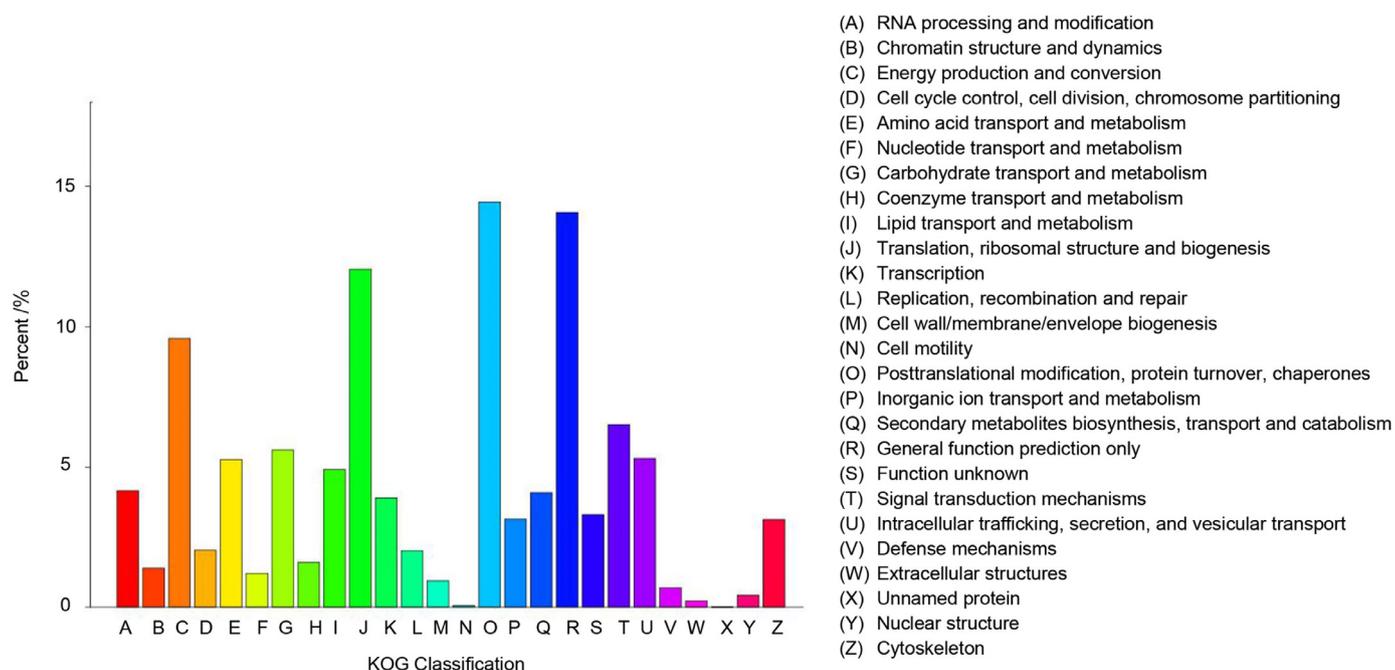


Fig. 4. Histogram presentation of KOG function classification of *L. casuarinoides* unigenes. All unigenes were subjected to a search against the KOG database for the functional prediction and classification. In total, 50,534 sequences were assigned to 25 KOG categories.

27,006 unigenes were found to have significant matches in the KEGG database, and the corresponding EC numbers were obtained from BLASTX alignments. Unigenes were assigned to 131 KEGG pathways that were divided into six main categories. Metabolic pathways had the largest number of unigenes (13,034, 48.26%), followed by “genetic information processing” (5902, 21.85%). The subcategories with most representations were “translation” (2939 unigenes, 10.88%), “carbohydrate metabolism” (2864, 10.6%), and “overview” (2449, 9.06%). (Fig. 5A).

In the metabolic pathway category, the most represented subcategories were carbohydrate metabolism (2864 unigenes), followed by metabolism of energy, lipids (1167), other amino acids (729), cofactors and vitamins (654), biosynthesis of others secondary metabolites (582), nucleotides (512), terpenoids and polyketides (410), as well as glycan biosynthesis and metabolism (184) (Fig. 5B).

In the secondary metabolism category, 1324 unigenes were classified into 21 subcategories (including others secondary metabolites and terpenoids and polyketides), and most of them were mapped to phenylpropanoid biosynthesis, carotenoid biosynthesis, terpenoid backbone biosynthesis (Fig. 5C). Interestingly, Tropane, piperidine and pyridine alkaloid biosynthesis involved 78 unigenes and ranked top six (Fig. 5C). These unigenes might serve as good candidates for identifying genes that participate in the lycopodium alkaloids biosynthesis pathway.

Additionally, 5902 unigenes were sorted to the genetic information processing involving translation, folding, sorting replication and repair, and 3637 were classified into transport and catabolism, signal transduction, membrane transport. These results demonstrate the power of RNA sequencing in identifying novel genes of non-model medicinal plants, and these annotations lay a foundation for investigating specific processes, functions and pathways involved in secondary metabolites of *L. casuarinoides*.

3.5. Candidate genes involved in lycodine-type alkaloids biosynthesis

Next, we focused on the discovery of genes involved in the lycodine-type alkaloids biosynthesis of *L. casuarinoides*, which

could be similar to those of *H. serrata*. Previous studies (Ma & Gang, 2004; Yang et al., 2017) showed that the lycodine-type lycopodium alkaloids biosynthesis in *H. serrata* is derived from the common pathway including the formation of the precursors, ring closure for the tetracyclic skeleton, and the subsequent modification and oxidation (Fig. 6). This figure is better to be the proposed biosynthesis pathway of *L. casuarinoides*. It is well known that *L. casuarinoides* is rich in lycodine-type lycopodium alkaloids (Tang et al., 2013; Zhang et al., 2014). The only differences between *H. serrata* and *L. casuarinoides* are types of oxidation leading to the formation of different products in the later steps of lycopodium alkaloids biosynthesis. Based on the KEGG pathway annotation, we identified all possible genes involved in the biosynthesis of lycopodium alkaloids in *L. casuarinoides*, including, LDC (lysine decarboxylase), PAO (primary amine oxidase), and malonyl-CoA decarboxylase, as listed in Table 2.

CYPs could determine the types of lycopodium alkaloids through oxidative modifications at the late stage of the alkaloid biosynthesis (Grabherr et al., 2011; Yang et al., 2017). Thus, the CYPs are also identified from the transcriptomic data. Based on the Swiss-Prot protein database, 64 CYP unigenes, belonging to 32 CYP families, were identified (Table 3). Among them, CYP52 (seven unigenes) are the most abundant, accounting for 10.94% of all CYP sequences, which is followed by CYP71 (6) and CYP734 (4), and the remaining CYPs account for 73.44% of all CYP sequences.

3.6. Transcription factors (TFs)

TFs play a critical role in regulating the secondary metabolism, as they can regulate the expression of related genes at the transcription level to control the flux of secondary metabolites. In order to identify the TFs in the transcriptome datasets of *L. casuarinoides*, unigene sequences were searched against the Plant Transcription Factor Database (PlnTFDB) based on the BLASTX. Totally 827 unigenes were shown to belong to 48 plant known transcription factor families. Among them, 83, 81, 72, 63, 52 and 50 unigenes were annotated as the C3H, Homeobox, MYB, bHLH, ERF, WRKY families, respectively (Fig. 7).

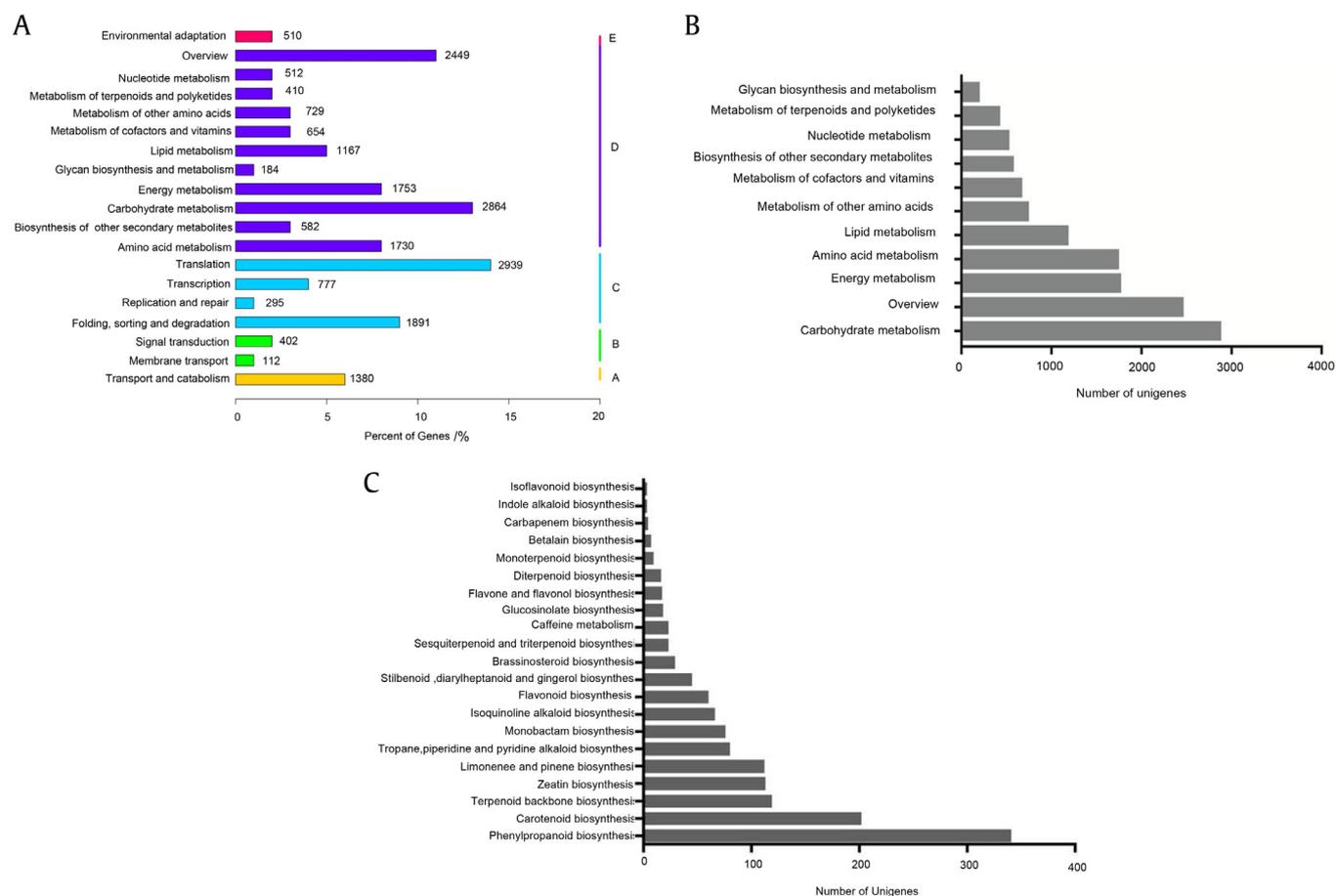


Fig. 5. KEGG analysis. (A) Pathway assignment based on KEGG database (In the right half, A, transport and catabolism; B, environmental information processing; C, genetic information processing; D, metabolism; E, environmental adaptation). (B) Classification based on metabolism categories; (C) Classification based on secondary metabolism categories. Right y-axis indicates specific category of genes in main category.

3.7. SSR detection and frequencies

Using the MISA software, 13,352 microsatellites (SSRs) were identified in 124,524 unigenes. Of all the SSR-containing unigenes, 1717 contained more than one SSR, and 727 SSRs were present in the compound form. On average, we found one SSR per 5.6 Kb length of unigenes (Table S2). There were 6118 (45.82%) dinucleotide motifs, 2792 (20.91%) trinucleotide motifs, 239 (1.79%) tetranucleotide motifs, 26 (0.19%) pentanucleotide motifs and 10 (0.07%) hexanucleotide motifs (Table 4). The length of SSRs was also analyzed; The majority were between 18 bp to 21 bp. SSRs with six tandem repeats (1956, 14.56%) were the most common, followed by five tandem repeats (1718, 12.87%), nine tandem repeats (1524, 11.43%), and seven tandem repeats (1494, 11.20%) (Table 4). The details of SSRs derived from all unigenes are shown in Data Sheet 2 (Supplementary materials). The most abundant repeat type was AG/CT (50.1%), followed by AC/GT (10.52%), AAG/CTT (8.98%), and AGC/CTG (6.6%). To verify the SSR, 25 SSRs were randomly selected and 25 pairs of PCR primers were designed using the software Primer 3. It turned out that 12 pairs of primers were applicable in the PCR amplification of the corresponding SSRs (Fig. S6), and the size of the PCR products was consistent with expected results (Table S3).

4. Discussion

Lycopodium alkaloids are major secondary metabolites and important medicinal substances of *L. casuarinoides*. Understanding

their biosynthesis pathway is important for the high-yield production in the wild species, but the basic genomic and transcriptional information of *Lycopodium* is scarce. In the last decade, high-throughput sequencing technologies help us to find out novel genes involved in secondary metabolite formation and to understand the biosynthesis pathway of various medicinal plants (Wei, Xiao, Hayward, & Fu, 2013). The *de novo* transcriptome assembly and the following bioinformatics analysis have become the efficient approach to infer the lycopodium alkaloids biosynthesis pathway in traditional medicinal plants such as *H. serrata* (Luo et al., 2010b; Yang et al., 2017) and *P. carinatus* (Luo et al., 2010b). In this study, we performed the RNA-Seq to profile the *L. casuarinoides* transcriptome. We obtained 16.43 Gb of data with 109,546,778 clean sequencing reads, and 61,304 unigenes (49.2% of the assem-

Table 2
Unigenes involved in biosynthesis of lycodine-type alkaloids in *L. casuarinoides*.

Gene names	EC no.	No. of Unigene
Lysine biosynthesis-related regulatory protein	NA	170
LDC, lysine decarboxylase	4.1.1.18	10
TYDC, tyrosine decarboxylase	4.1.1.25	1
Arginine decarboxylase	4.1.1.19	4
Pyridoxal 5'-phosphate synthase	4.3.3.6	19
Cadaverine /lysine antiporter	NA	1
PAO, primary amine oxidase	1.4.3.21	7
Malonyl-CoA decarboxylase	4.1.1.94	1
Methylmalonyl-CoA mutase	5.4.99.2	2
Malonyl-CoA-acyl carrier protein transacylase	2.3.1.39	2

Note: NA, not applicable.

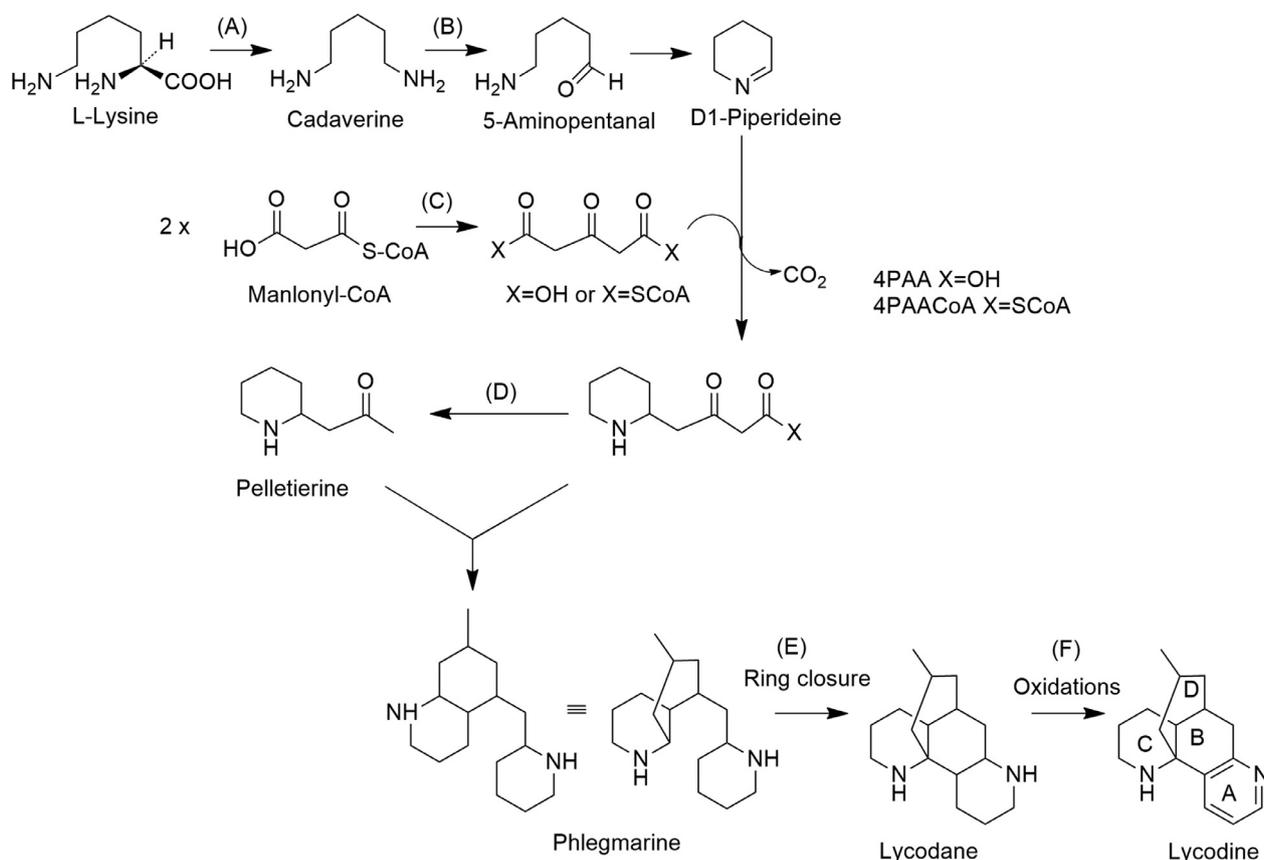


Fig. 6. Putative pathways for lycodine-type alkaloid biosynthesis in *L. casuarinoides*. Enzymes found in this study were as follows: (A) LDC, lysine decarboxylase; (B) CAO, copper amine oxidase; (C) PKS, polyketide synthase; (D) unknown enzyme; (E) and (F) CYPs, cytochrome P450.

bled unigenes) were successfully annotated with the public protein databases.

What are their bioactivity and therapeutic efficacy? For the deep understanding of the relationship between lycopodium alkaloids and the corresponding biosynthesis genes of *Lycopodium*, 43,080 and 27,006 unigenes were assigned to GO and KEGG categories respectively. We found that 1324 unigenes may be related to the various types of alkaloids, e.g., tropane, piperidine, and pyridine alkaloids. More importantly, several critical biosynthesis genes

encoding LDC, PAO, and malonyl-CoA decarboxylase were successfully identified in transcriptome datasets. It seems that the biosynthesis of lycopodium alkaloids in *L. casuarinoides* might be similar to that in *H. serrata*, and both plants share similar early enzymatic steps in their biosynthetic pathways. Given the absence of HupaA and the presence of huperzine B/C in *L. casuarinoides* (Zhang et al., 2014), we speculate that the alkaloid types may depend on later steps of biosynthesis, e.g., the modification and cyclization of alkaloid ring structure (Fig. 6). This figure is better to be the pro-

Table 3
Summary of CYP family identified in *L. casuarinoides* transcriptome dataset.

Gene names	No. of unigene	Type	Percentage /%	Gene name	No. of unigene	Type	Percentage /%
CYP1	1	A1	1.56	CYP84	1	A4	1.56
CYP2	1	D6	1.56	CYP85	2	A1/ A2	3.13
CYP3	2	A2/A8	3.13	CYP86	2	A8/ B1	3.13
CYP4	2	G15/V2	3.13	CYP89	1	A2	1.56
CYP6	2	B6/K1/D5	3.13	CYP90	1	A1	1.56
CYP9	1	E2	1.56	CYP93	1	A3	1.56
CYP11	1	B1	1.56	CYP94	1	B3	1.56
CYP52	7	A11/A12/ A13/ A2/ A3/A9/E1	10.94	CYP98	3	A1/ A2/ A3	4.69
CYP55	2	A2,A3	3.13	CYP318	1	A1	1.56
CYP61	2	NA	3.13	CYP703	1	A2	1.56
CYP71	6	A1/A22/ B12/B13/B24/ B28	9.38	CYP704	2	B1/C1	3.13
CYP72	3	A11/ A13/ A15	4.69	CYP710	2	A1/ A2	3.13
CYP76	2	C2/C3	3.13	CYP714	1	B2	1.56
CYP77	1	A1	1.56	CYP716	2	B1/ B2	3.13
CYP78	2	A4/ A5	3.13	CYP734	4	A2/ A4/ A6	6.25
CYP81	3	D1/D11/F1	4.69	CYP750	1	A1	1.56

Note: CYP denotes Cytochrome P450; No. of genes means the number of unique putative transcripts with homology to cytochrome P450s.
 †Type refers to abbreviation of the identified CYP450 genes; NA, not applicable.

Table 4
Distribution of identified SSRs.

Motif	Repeat numbers								Total	%
	5	6	7	8	9	10	11	12		
Di-	0	1148	992	1157	1521	1063	229	8	6118	45.82
Tri-	1486	768	502	33	2	–	–	1	2792	20.91
Tetra-	202	35	0	2	0	–	–	0	239	1.79
Penta-	22	3	0	0	1	–	–	0	26	0.19
Hexa-	8	2	0	0	0	–	–	0	10	0.07
Total	1718	1956	1494	1192	1524	1063	229	9	–	–
%	12.87	14.65	11.20	8.93	11.43	7.96	1.72	0.067	–	–

Note: “–” denotes no data.

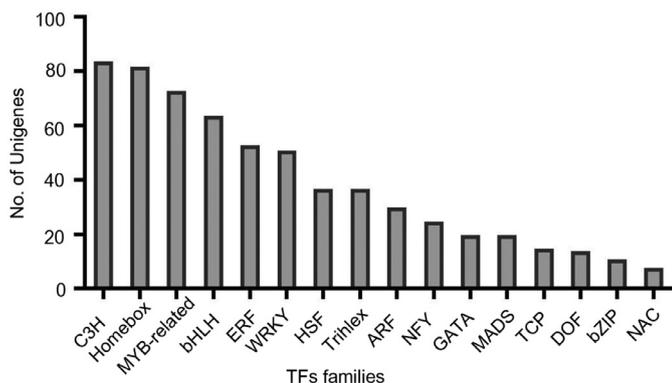


Fig. 7. Major TF families identified in *L. casuarinoides* transcriptome dataset. Type and number of TFs identified in *L. casuarinoides* were shown. C3H, Homeobox, MYB, bHLH and ERF proteins were the most abundant.

posed biosynthesis pathway of *L. casuarinoides*. Intriguingly, some novel genes possibly involved in the biosynthesis of lycopodium alkaloids are also annotated, such as a unigene encoding the cadaverine/lysine antiporter, and two unigenes encoding malonyl-CoA-acyl carrier protein transacylase (Table 2). Moreover, we found 64 CYPs in the *L. casuarinoides* transcriptome dataset, such as CYP52 and CYP71. However, two key classes of CYPs that function in the modification of lycopodium alkaloids scaffold (Yang et al., 2017), e.g., BBE (berberine bridge enzyme, encoded by CYP719) and SLS (secologanin synthase, encoded by CYP72A1), were discovered in the *H. serrata* transcriptome (Yang et al., 2017). However, neither of them was identified in *L. casuarinoides*. Therefore, we infer that other isoforms of CYPs might participate in regulating the late two stages of ring formation and oxidation (Fig. 6). Owing to the limited samples, the composition and content of lycopodium alkaloids were not investigated at the moment, more investigations are warranted for elucidating the link between the gene expressions and the metabolic phenotypes.

Several TF families are essential in the regulation of alkaloid biosynthesis genes in plants. For example, in opium poppy (*Papaver somniferum* L.), the C3H type TF Ps175C3H was identified to be involved in the papaverine biosynthesis (Agarwal et al., 2016). In *Coptis japonicas*, CjbHLH1 could regulate the isoquinoline alkaloid biosynthesis (Yamada et al., 2011). In the present study, we detected 827 *L. casuarinoides* unigenes representing homologs belonging to various TF families including C3H, Homeobox, MYB, bHLH, ERF and WRKY, etc. The top three TF families in our datasets are C3H, Homeobox and MYB, which is comparable to those of two closely related species *H. serrata* and *P. carinatus* (Luo et al., 2010a). Their exact functions are worth further study.

As is well known, SSR molecular markers are useful for the study of population structure, genetic diversity and genetic linkage mapping (Jiang et al., 2013). In this study, 13,352 potential SSRs were identified from 124,524 unigenes of *L. casuarinoides*. Except

for the mononucleotide repeats, the dinucleotide repeats (8077 SSRs, 50.1%) were the most abundant repeat type, followed by trinucleotide repeats (2080, 15.58%). AG/CT (6689, 50.1%) was the most abundant dinucleotide repeats in *L. casuarinoides*, so it is in *H. serrata* and *P. carinatus* (Luo et al., 2010b).

5. Conclusion

In the present study, we, for the first time, performed the *de novo* transcriptome sequencing analysis of *L. casuarinoides* aerial tissues on the Illumina platform. More than 16.43 Gb of sequencing data were generated and assembled into 61,304 unigenes. We identified a large number of candidate genes potentially involved in secondary metabolic pathway, including genes related to the biosynthesis of lycodine-type alkaloids. LDC, PAO and a series of genes related to the regulation and modification of lycopodium alkaloids biosynthesis, e.g., TFs and CYPs, were identified. In addition, a large number of SSRs were detected in the transcriptome dataset. This dataset might provide useful information about the key biosynthetic genes of *L. casuarinoides*. Further studies are needed to elucidate the CYPs involved in the ring formation and oxidative modification of the biosynthesis of huperzine B/C. This preliminary study provides valuable resources for bioengineering and synthetic biology studies of the lycopodium alkaloids.

Declaration of Competing Interest

The authors declare no conflict of interests.

Acknowledgments

This study was supported by Jiangxi Foreign Science and Technology Cooperation Major Projects (Grant no. 20151BDH80020), Science and Technology Project of Jiangxi Education Department (Grant no. GJJ180846) and Jiangxi Postdoctoral Fund Project (Grant no. JX2016RC40). We thank Mr. Wei Gao, Ms. Tang Wei, and Ms. Jane Zhen for their good advice of the assays.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.chmed.2020.03.001.

References

- Agarwal, P., Pathak, S., Lakhwani, D., Gupta, P., Asif, M. H., & Trivedi, P. K. (2016). Comparative analysis of transcription factor gene families from *Papaver somniferum*: identification of regulatory factors involved in benzyloisoquinoline alkaloid biosynthesis. *Protoplasma*, 253, 857–871.
- Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, 21, 3674–3676.
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, 29, 644–652.

- Hirasawa, Y., Kato, E., Kobayashi, J., Kawahara, N., Goda, Y., Shiro, M., et al. (2008). Lycoparins A-C, new alkaloids from *Lycopodium casuarinoides* inhibiting acetylcholinesterase. *Bioorganic & Medicinal Chemistry*, 16, 6167–6171.
- Iseli, C., Jongeneel, C. V., & Bucher, P. (1999). ESTScan: A program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. In *Proceedings of the International Conference on Intelligent Systems for Molecular Biology* (pp. 138–148). Proceedings / ... International Conference on Intelligent Systems for Molecular Biology.
- Jiang, F., Li, W., Huang, Y., Chen, Y., Jin, B., Chen, N., et al. (2013). Antioxidant, antityrosinase and antitumor activity comparison: The potential utilization of fibrous root part of *Bletilla striata* (Thunb.) Reichb.f. *PLoS ONE*, 8(2), e58004.
- Jiang, W. W., Liu, F., Gao, X., He, J., Cheng, X., & Peng, L. Y. (2014). Huperzines A-E, lycopodium alkaloids from *Huperzia serrata*. *Fitoterapia*, 99, 72–77.
- Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Research*, 36, D480–D484.
- Luo, H., Li, Y., Sun, C., Wu, Q., Song, J., Sun, Y., et al. (2010a). Comparison of 454-ESTs from *Huperzia serrata* and *Phlegmariurus carinatus* reveals putative genes involved in lycopodium alkaloid biosynthesis and developmental regulation. *BMC Plant Biology*, 10, 209.
- Luo, H., Sun, C., Li, Y., Wu, Q., Song, J., Wang, D., et al. (2010b). Analysis of expressed sequence tags from the *Huperzia serrata* leaf for gene discovery in the areas of secondary metabolite biosynthesis and development regulation. *Physiologia Plantarum*, 139, 1–12.
- Ma, X., & Gang, D. R. (2004). The lycopodium alkaloids. *Natural Product Reports*, 21, 752–772.
- Ma, X. Q., Jiang, S. H., & Zhu, D. Y. (1998). Alkaloid patterns in *Huperzia* and some related genera of *Lycopodiaceae sensu lato* occurring in China and their contribution to classification. *Biochemical Systematics and Ecology*, 26, 723–728.
- Pan, K., Xia, X., Guo, W. H., & Kong, L. Y. (2015). Suppressive effects of total alkaloids of *Lycopodium casuarinoides* on adjuvant-induced arthritis in rats. *Journal of Ethnopharmacology*, 159, 17–22.
- Pertea, G., Huang, X., Liang, F., Antonescu, V., Sultana, R., Karamycheva, S., et al. (2003). TIGR gene indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinformatics*, 19, 651–652.
- Sharma, S., & Shrivastava, N. (2016). Renaissance in phytomedicines: promising implications of NGS technologies. *Planta*, 244, 19–38.
- Tang, Y., Fu, Y., Xiong, J., Li, M., Ma, G. L., & Yang, G. X. (2013). Casuarinines A-J, lycodine-type alkaloids from *Lycopodium casuarinoides*. *Journal of Natural Products*, 76, 1475–1484.
- Wei, L., Xiao, M., Hayward, A., & Fu, D. (2013). Applications and challenges of next-generation sequencing in *Brassica* species. *Planta*, 238, 1005–1024.
- Yamada, Y., Kokabu, Y., Chaki, K., Yoshimoto, T., Ohgaki, M., Yoshida, S., et al. (2011). Isoquinoline alkaloid biosynthesis is regulated by a unique bHLH-type transcription factor in *Coptis japonica*. *Plant and Cell Physiology*, 52, 1131–1141.
- Yang, M., You, W., Wu, S., Fan, Z., Xu, B., Zhu, M., et al. (2017). Global transcriptome analysis of *Huperzia serrata* and identification of critical genes involved in the biosynthesis of huperzine A. *BMC Genomics*, 18, 245.
- Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., et al. (2006). WEGO: A web tool for plotting GO annotations. *Nucleic Acids Research*, 34, W293–W297.
- Zhang, D. B., Chen, J. J., Song, Q. Y., Zhang, L., & Gao, K. (2014). Lycodine-type alkaloids from *Lycopodium casuarinoides* and their acetylcholinesterase inhibitory activity. *Molecules*, 19, 9999–10010.