



Full-length transcriptome analysis of *Adiantum flabellulatum* gametophyte

Zeping Cai^{1,*}, Zhenyu Xie^{1,*}, Luyao Huang¹, Zixuan Wang¹, Min Pan¹, Xudong Yu¹, Shitao Xu² and Jiajia Luo³

¹ Key Laboratory of Genetics and Germplasm Innovation of Tropical Special Forest Trees and Ornamental Plants, Ministry of Education, College of Forestry, Hainan University, Haikou, Hainan, China

² College of Horticulture, Hainan University, Haikou, Hainan, China

³ Tropical Crops Genetic Resources Institute, Chinese Academy of Tropical Agricultural Sciences, Haikou, Hainan, China

*These authors contributed equally to this work.

ABSTRACT

Ferns are important components of plant communities on earth, but their genomes are generally very large, with many redundant genes, making whole genome sequencing of ferns prohibitively expensive and time-consuming. This means there is a significant lack of fern reference genomes, making molecular biology research difficult. The gametophytes of ferns can survive independently, are responsible for sexual reproduction and the feeding of young sporophytes, and play an important role in the alternation of generations. For this study, we selected *Adiantum flabellulatum* as it has both ornamental and medicinal value and is also an indicator plant of acidic soil. The full-length transcriptome sequencing of its gametophytes was carried out using PacBio three-generation sequencing technology. A total of 354,228 transcripts were obtained, and 231,705 coding sequences (CDSs) were predicted, including 5,749 transcription factors (TFs), 2,214 transcription regulators (TRs) and 4,950 protein kinases (PKs). The transcripts annotated by non-redundant protein sequence database (NR), Kyoto encyclopedia of genes and genomes (KEGG), eukaryotic ortholog groups (KOG), Swissprot, protein family (Pfam), nucleotide sequence database (NT) and gene ontology (GO) were 251,501, 197,474, 193,630, 194,639, 195,956, 113,069 and 197,883, respectively. In addition, 138,995 simple sequence repeats (SSRs) and 111,793 long non-coding RNAs (lncRNAs) were obtained. We selected nine chlorophyll synthase genes for qRT-PCR, and the results showed that the full-length transcript sequences and the annotation information were reliable. This study can provide a reference gene set for subsequent gene expression quantification.

Subjects Bioinformatics, Molecular Biology, Plant Science

Keywords Ferns, Three-generation sequencing, Functional annotation, Reference gene set

INTRODUCTION

Ferns are an ancient group of land plants. Their origins can be traced back to 380 million years ago (*Schneider et al., 2004*). Ferns are the second largest group of vascular plants (*Sessa et al., 2014*). At present, more than 11,000 species have been recorded (*The Pteridophyte Phylogeny Group, 2016*). They are important components of plant communities on Earth.

Submitted 24 September 2021

Accepted 16 February 2022

Published 9 March 2022

Corresponding author

Jiajia Luo, luojiajia09skj@163.com

Academic editor

Vladimir Uversky

Additional Information and
Declarations can be found on
page 16

DOI 10.7717/peerj.13079

© Copyright
2022 Cai et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

The genomes of ferns are generally large, with an average size of 12 Gb, and the largest can even reach 148 Gb ([Hidalgo et al., 2017a](#); [Hidalgo et al., 2017b](#)). *Ceratopteris thalictroides* is a model plant of ferns with a genome size of 11.25 Gb, which is more than 80 times the size of the genome of *Arabidopsis thaliana* (135 Mb; [Marchant et al., 2019](#)). *Adiantum flabellulatum* belongs to the genus *Adiantum* of Pteridaceae, whose genome size is unknown. The smaller genomes of *Adiantum caudatum* and *Anogramma leptophylla* in Pteridaceae are 3.78 Gb and 3.0 Gb, respectively ([Kuo & Li, 2019](#)). Because the genes of a fern are highly redundant, whole genome sequencing of ferns is both expensive and time-consuming. The genomes of *Azolla filiculoides* and *Salvinia cucullata* are relatively small at 0.75 Gb and 0.26 Gb, respectively ([Li et al., 2018](#)). They are currently the only two ferns with whole genome sequences. Due to the lack of reference genomes, the molecular biology research of ferns has somewhat stagnated.

The gametophytes of ferns are very small, but they can survive independently. They are responsible for sexual reproduction, feeding young sporophytes, and are indispensable in the process of alternation of generations. *A. flabellulatum* is mainly distributed in China, Japan, and Vietnam. [Cao, Huang & Wang \(2010\)](#) observed the gametophyte development process of *A. flabellulatum* by microscope. [Wang et al. \(2018\)](#) found that plant hormone brassinosteroids (BRs) could regulate the gametophyte development of *A. flabellulatum*.

For plants without reference genomes, full-length transcripts can effectively provide a reference gene set for gene expression quantification. [Jia et al. \(2020\)](#), for example, used PacBio single-molecule real-time (SMRT) sequencing technology to analyze the full-length transcriptome of *Rhododendron lapponicum*, and obtained a large number of full-length transcripts, which facilitated gene expression quantification. [Wu et al. \(2020\)](#) analyzed the sequencing data of the full-length transcriptome of endangered species *Populus wulianensis*, which provided a scientific foundation for an in-depth study of its gene functions. In this study, the full-length transcripts of *A. flabellulatum* gametophytes were obtained using PacBio three-generation sequencing technology, which can provide a reference gene set for subsequent gene expression quantification and is of great significance to help further the research of gametophyte development in ferns.

MATERIALS & METHODS

Plant materials

The spores of *A. flabellulatum* were collected from the rubber forest in Ma'an mountain, Danzhou City, Hainan Province, China (109.519°E, 19.501°N). The mature spores were sterilized and inoculated in 1/4 MS medium to form young gametophytes after germination. Young gametophytes with diameters of 1–2 mm were selected and transferred into new 1/4 MS medium. These young gametophytes were then placed in illuminations of 0, 10, 500, 1,000 and 10,000 lx, respectively, and cultured at 25 °C for 12 days. Three biological replicates were set in the above 5 illuminations, respectively, with a total of 15 samples. Approximately 200 µg of gametophytes from each sample were selected for RNA extraction.

RNA extraction, Library construction and Sequencing

Total RNA (more than 40 ng/ μ L) was extracted from the above 15 samples using the CTAB method, and the mRNAs were purified and mixed to construct a cDNA library (Yu *et al.*, 2021). Raw reads were obtained on the PacBio Sequel platform. After reads of insert (ROI) were generated from raw reads, clustering and correction (Chin *et al.*, 2013) were performed using the Quiver algorithm to obtain high-quality full-length consensus reads and then the transcripts were obtained after redundancy removal (using the Cd-hit program, Li & Godzik, 2006). RNA extraction, library construction and sequencing were performed by the Beijing Genomics Institute (BGI).

Coding sequences prediction

The Transdecoder software (v3.0.1, Haas *et al.*, 2013) was used to identify the candidate coding regions in the transcripts. The longest open reading frames were extracted, and the coding sequences (CDSs) were predicted using the SwissProt (<http://ftp.ebi.ac.uk/pub/databases/swissprot>, Boutet *et al.*, 2007) and protein families (Pfam, <http://pfam.xfam.org>, Finn *et al.*, 2016) databases.

Simple sequence repeats detection

The MISA software (v1.0) (Thiel *et al.*, 2003) was used to detect the simple sequence repeats (SSRs). The detection parameters were mono-nucleotide, di-nucleotide, tri-nucleotide, tetra-nucleotide, penta-nucleotide and hexa-nucleotide repeats at least 12, 6, 5, 5, 4 and 4 times, respectively. The Primer3 software (v2.2.2) (Untergasser *et al.*, 2012) was used to design primers for the SSRs.

Functional annotation

BLASTN was used for NT (<https://blast.ncbi.nlm.nih.gov/Blast>) annotations; Blastx (BLAST, v2.2.23, Altschul *et al.*, 1990) or Diamond (v0.8.31) (Buchfink, Xie & Huson, 2015) were used for NR (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), KEGG (<http://www.genome.jp/kegg>), KOG (<http://www.ncbi.nlm.nih.gov/KOG>) and SwissProt annotations; and Blast2GO (v2.5.0) (Conesa *et al.*, 2005) was used for GO (<http://geneontology.org>) annotations. After translating CDSs into protein sequences, the protein sequences were uploaded to iTAK (<http://itak.feilab.net/cgi-bin/itak/index.cgi>) for the prediction of transcription factors (TFs), transcription regulators (TRs) and protein kinases (PKs).

Long non-coding RNAs prediction

The Coding Potential Calculator (CPC) (v0.9-r2) (Kong *et al.*, 2007), txCdsPredict (<http://hgdownload.soe.ucsc.edu/admin/fksrc.zip>) (Zweig *et al.*, 2008) and the Coding-Non-Coding Index (CNCI) (<https://github.com/www-bioinfo-org/CNCI>) (Sun *et al.*, 2013) were used to score transcripts other than CDSs, and the scoring thresholds (CPC_threshold = 0, txCdsPredict_threshold = 500, CNCI_threshold = 0) were set to determine long non-coding RNAs (lncRNAs). Pfam was used to predict the protein-coding potential of a transcript. The transcript that could not be predicted in Pfam is potential lncRNA. If the results with three of the above four methods were consistent, the transcript was finally confirmed as lncRNA.

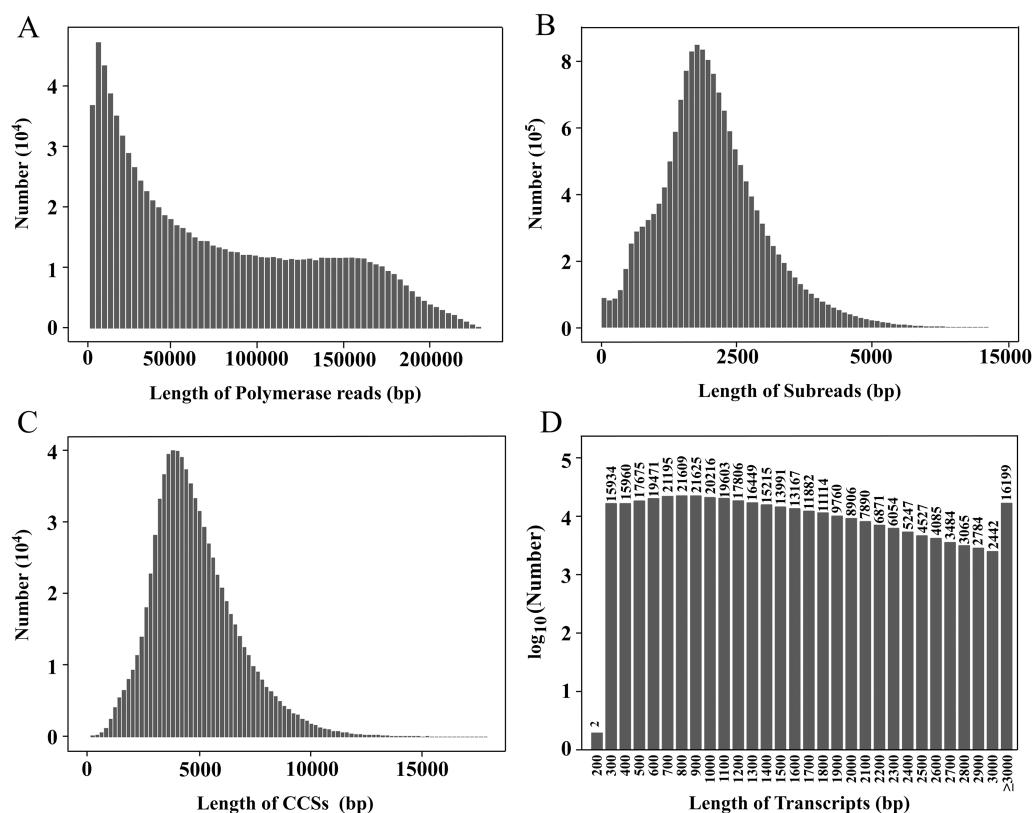


Figure 1 Length distributions of PacBio SMRT sequencing. (A–D) The length distributions of polymerase reads (A), subreads (B), CCSs (C), and transcripts (D), respectively.

Full-size [DOI: 10.7717/peerj.13079/fig-1](https://doi.org/10.7717/peerj.13079/fig-1)

Quantitative real-time polymerase chain reaction

In order to verify the reliability of the full-length transcript sequences and the annotation information, we selected nine chlorophyll synthase genes and used the method applied by [Cai et al. \(2021\)](#) for qRT-PCR ([Table S1](#)). Each of the four treatments (0 lx, 10 lx, 500 lx and 10,000 lx) were detected with three biological replicates for a total of 12 samples. Three parallel tests were performed for each sample, and the isoform_55882 (*Ubiquitin C*) was used as a reference gene. Relative quantification was performed using the $2^{-\Delta\Delta Ct}$ method ([Pfaffl, 2001](#); [Cai et al., 2021](#)).

RESULTS

PacBio SMRT sequencing data

In this study, a PacBio ISOSeq library was established, and the full-length transcriptome sequencing was performed on the PacBio Sequel platform. The total base of polymerase reads was 69.87 Gb, and the total number of reads was 923,898. The average length, the longest sequence and the N50 of polymerase reads were 75,625.26 bp, 303,495 bp and 129,782 bp, respectively ([Fig. 1A](#)). After removing the adapters, 16,221,787 subreads were obtained, with the total base of 68.72 Gb. The average length, the longest sequence and the N50 of subreads were 4,236.18 bp, 218,714 bp and 4,762 bp, respectively ([Fig. 1B](#)).

Table 1 PacBio SMRT sequencing data statistics.

Item	Total reads	Total base (GB)	Max length (bp)	Mean length (bp)	N50 (bp)
Polymerase reads	923,898	69.87	303,495	75,625.26	129,782
Subreads	16,221,787	68.72	218,714	4,236.18	4,762
Circular consensus sequences (CCSs)	809,367	3.99	62,285	4,934	4,521
Full-length non-chimeric (FLNC) reads	2,762,613	2.76	41,488	998	873
Consensus reads	2,608,705	2.59	41,488	994	871
Transcripts	354,228	0.46	41,488	1,308.60	1,667

Table 2 Quality metrics of predicted CDSs.

Total number	Total length (bp)	N50 (bp)	N90 (bp)	Max length (bp)	Min length (bp)	GC (%)
231,705	194,628,012	1,056	420	20,595	297	47.89%

After subreads from the same circular molecule were combined into circular consensus sequences (CCSs), a total of 809,367 CCSs were obtained. The average length and average quality were 4,934 bp and 0.97, respectively (Fig. 1C). The CCSs were split into full-length, non-full-length, chimeric and non-chimeric sequences according to whether 5' primer, 3' primer and ploy(A) tail were detected. There were 2,762,613 full-length non-chimeric (FLNC) sequences, with an average length of 998 bp and an average quality of 0.98 (Fig. S1).

The FLNC sequences from the same copy were grouped into clusters, and the Arrow algorithm (Hong *et al.*, 2020) was used to correct the error. After removing the low-quality sequences, 2,608,705 consensus reads were obtained with an average quality of 0.99 (Fig. S2). After removing redundancy, 354,228 transcripts were finally obtained (Table 1, Fig. 1D). The CCS and transcript data have been uploaded to the NCBI Sequence Read Archive (SRA) database (<https://www.ncbi.nlm.nih.gov/sra>), and the accession numbers are PRJNA774117 and PRJNA733457, respectively.

CDS prediction

CDS prediction was performed on transcripts, and a total of 231,705 CDSs were obtained. The total length was 194,628,012 bp, and the N50 was 1,056 bp (Table 2). There were 164,535 CDSs with a length of 100–1,000 bp, accounting for 71.0% of the total. There were 57,009 CDSs of 1,000–2,000 bp, 8,722 CDSs of 2,000–3,000 bp and 1,439 CDSs of \geq 3,000 bp, accounting for 24.6%, 3.8% and 0.6% of the total, respectively (SRA database; Fig. 2A). The CDSs have been uploaded to the SRA database, and the accession number is PRJNA777740.

SSR detection

Through SSR detection of 354,228 transcripts, 138,995 SSRs were obtained from 81,304 transcripts, with a total length of 3,875,607 bp (Table S2). Of these, 50,482 transcripts contained one SSR locus and 30,822 transcripts contained two or more SSR loci (Table 3). Among SSR repeat types, di-nucleotide repeats (90,741) accounted for the highest proportion (65.28%). The second was the mono-nucleotide repeats (35,856) accounting for 25.8% of total repeats and the third was tri-nucleotide repeats (10,528) accounting

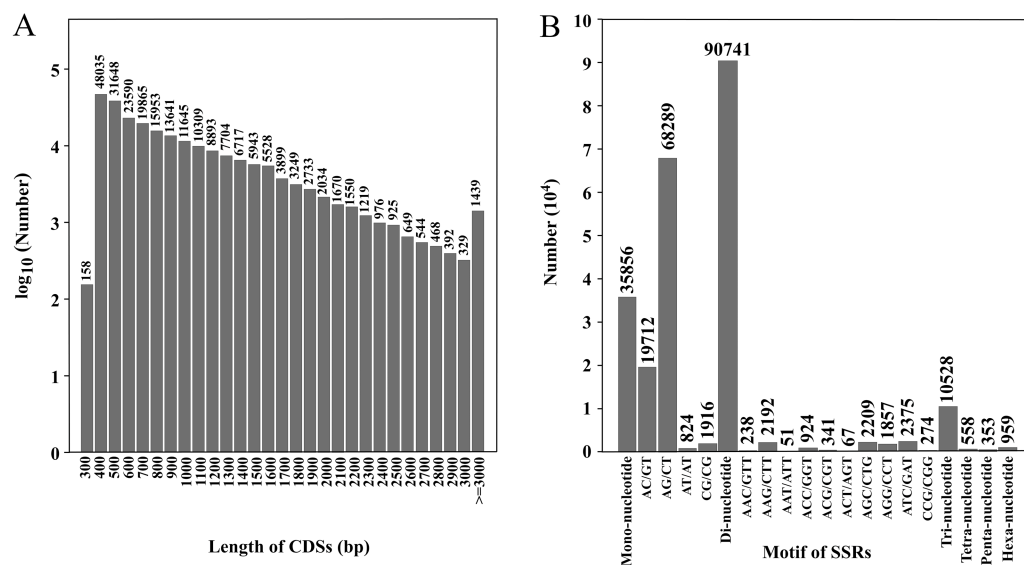


Figure 2 Length distribution of CDSs and number of various SSR motifs. (A) Number of CDSs in different length distributions; (B) number of various SSR motifs.

Full-size [DOI: 10.7717/peerj.13079/fig-2](https://doi.org/10.7717/peerj.13079/fig-2)

Table 3 Summary of SSRs identified in the transcripts.

Searching item	Number
Total number of transcripts examined	354,228
Total size of examined sequences (bp)	463,542,116
Total number of identified SSRs	138,995
Total length of SSRs	3,875,607
Number of SSR containing sequences	81,304
Number of sequences containing more than 1 SSR	30,822
Number of SSRs present in compound formation	41,368

for 7.56% of repeats. Tetra-nucleotide, penta-nucleotide and hexa-nucleotide repeats accounted for the minority at 0.40%, 0.25% and 0.69% of repeats, respectively. Forty-one thousand three hundred sixty-eight SSRs were present in compound formation. SSRs with six repeat units (19,697, 14.17%) were the most common, followed by seven repeat units (12,575, 9.05%), eight repeat units (9,806, 7.05%) and 12 repeat units (9,272, 6.67%) (Table 4). Among the SSRs with di-nucleotide repeats, AG/CT (68,289) was the most common type, and AT/AT (824) was the least. Among the SSRs with tri-nucleotide repeats, ATC/GAT (2,375) was the most common type and AAT/ATT (51) was the least common (Fig. 2B). Twenty-nine thousand five hundred eighty-nine pairs of primers were designed and screened by Primer3. The primer lengths were 18–28 bp, the GC contents were 20.00–72.22%, the annealing temperatures were 57–63 °C, and the product lengths were 80–276 bp (Table S3).

Table 4 Types of SSR repeat units detected.

Number of repeat units	Mono nucleotide	Di nucleotide	Tri nucleotide	Quad nucleotide	Penta nucleotide	Hexa nucleotide	Total	Percentage (%)
4	0	0	0	0	291	786	1,077	0.77
5	0	0	6,278	377	48	133	6,836	4.92
6	0	17,114	2,412	130	13	28	19,697	14.17
7	0	11,672	854	37	1	11	12,575	9.05
8	0	9,247	547	12	0	0	9,806	7.05
9	0	7,700	223	2	0	0	7,925	5.70
10	0	6,181	107	0	0	0	6,288	4.52
11	0	4,946	58	0	0	0	5,004	3.60
12	4,837	4,411	24	0	0	0	9,272	6.67
13	3,677	4,014	19	0	0	0	7,710	5.55
14	2,899	3,546	6	0	0	0	6,451	4.64
≥15	24,443	21,910	0	0	0	1	46,354	33.35

Table 5 Number of transcripts annotated by seven databases.

Item	Total	NR	KEGG	KOG	Swissprot	Pfma	NT	GO	Intersection	Overall
Number	354,228	251,501	197,474	193,630	194,639	195,956	113,069	197,883	55,94	277,198
Percentage	100%	71.00%	55.75%	54.66%	54.95%	55.32%	31.92%	55.86%	15.75%	78.25%

Notes.

Note: the “Intersection” expresses the number and proportion of transcripts annotated by 7 databases together; the “Overall” expresses the number and proportion of transcripts annotated by at least 1 database.

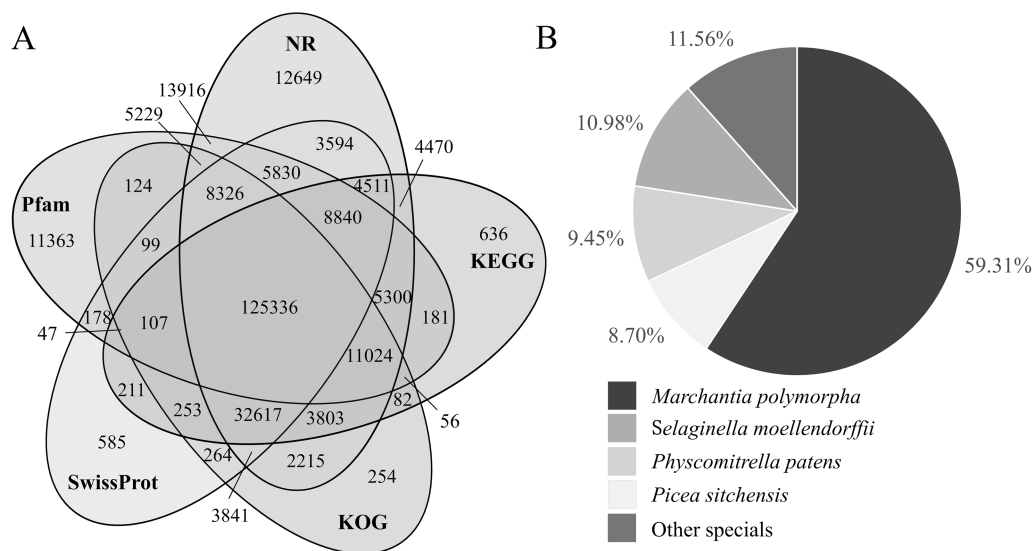


Figure 3 Venn diagram of transcript number distribution annotated by 5 databases and homologous species information in the NR database. (A) Venn diagram of transcripts annotated by NR, KEGG, KOG, SwissProt and Pfam databases. The numbers in the overlapping areas represent the numbers of shared genes. (B) The proportion of transcripts derived from homologous species and annotated by the NR database.

Full-size DOI: 10.7717/peerj.13079/fig-3

Functional annotation

Three hundred fifty-four thousand two hundred twenty-eight transcripts were annotated by seven databases. There were 251,501, 197,474, 193,630, 194,639, 195,956, 113,069 and 197,883 transcripts annotated by the NR, KEGG, KOG, SwissProt, Pfam, NT and GO databases (Table S4), respectively. Most of them were annotated by the NR database, accounting for 71.00%. The NT database annotated the least, accounting for 31.92% (Table 5). Fifty-five thousand seven hundred ninety-four transcripts were annotated by all the seven databases combined, accounting for 15.75% of the transcripts. There were 125,336 transcripts annotated by NR, KEGG, KOG, SwissProt and Pfam databases together, accounting for 35.38% (Fig. 3A).

NR annotation

The proportion of transcripts annotating to homologous species was counted from the NR annotation results. Among them, the proportions of transcripts annotating to *Marchantia polymorpha*, *Selaginella moellendorffii*, *Physcomitrella patens* and *Picea sitchensis*

were 59.31%, 10.98%, 9.45% and 8.70%, respectively, and the proportion of transcripts annotating to other known species was 11.56% (Table S5; Fig. 3B).

KEGG annotation

A total of 197,474 transcripts were annotated by the KEGG database, of which 112,839 were annotated to 137 pathways (Table S6). The KEGG pathways were divided into five categories: “cell processing,” “environmental information processing,” “genetic information processing,” “metabolism” and “organismal system.” For “metabolism,” “global and overview maps” had the greatest number of transcripts (55,321), accounting for 28.01%, followed by “carbohydrate metabolism,” which had 21,432 transcripts, accounting for 10.85% (Fig. 4A).

KOG annotation

There were 193,630 transcripts annotated by the KOG database, which could be divided into 25 groups according to KOG functional classification. Among them, the transcripts belonging to “general function prediction only” were the most common with a total of 48,175, accounting for 24.88%. There were 26,065 “signal transduction mechanisms” and 23,151 “posttranslational modifications protein turnover chaperones,” accounting for 13.46% and 11.96%, respectively (Fig. 4B).

GO annotation

A total of 197,883 transcripts were annotated by the GO database. GO domains can be divided into three categories: “biological processes,” “cell components” and “molecular functions.” A large number of transcripts in “biological processes” were mainly involved in the “cellular process” (94,069) and the “metabolic process” (81,073). The category “cell components” mainly consisted of transcripts involved in “cellular anatomical entity” (129,811) and “intracellular” (82,674). The “Cell components” category mainly consisted of transcripts involved in “binding” (95,413) and “catalytic activity” (95,176) (Fig. 4C).

Prediction of TFs, TRs and PKs

TFs play an important role in gene expression regulation. A total of 5,749 CDSs were predicted belonging to 64 TF families. Among them, the predicted sequences of C3H family were the most common (614), followed by bHLH (449), STAT (1) was the least common (Table 6; Table S7; Fig. 5). In addition, 2,214 sequences belonged to 24 TR families. Among them, the SET family had the most sequences (251), followed by the GNAT family (248), and the two families that had the least number of sequences were RB (4) and MED7 (4) (Table 6; Table S8; Fig. 6).

PKs regulate protein activity and amplify signals to induce a cellular response during plant signal transduction (Ardito *et al.*, 2017). A total of 4,950 CDSs were predicted belonging to the PK families. Among them, 833 sequences belonged to leucine rich-repeat receptor-like kinases (LRR-RLKs) families and 746 sequences belonged to receptor-like cytoplasmic kinase (RLCK) families, accounting for 16.83% and 15.07% of the total PKs, respectively (Table 6; Table S9; Fig. 7).

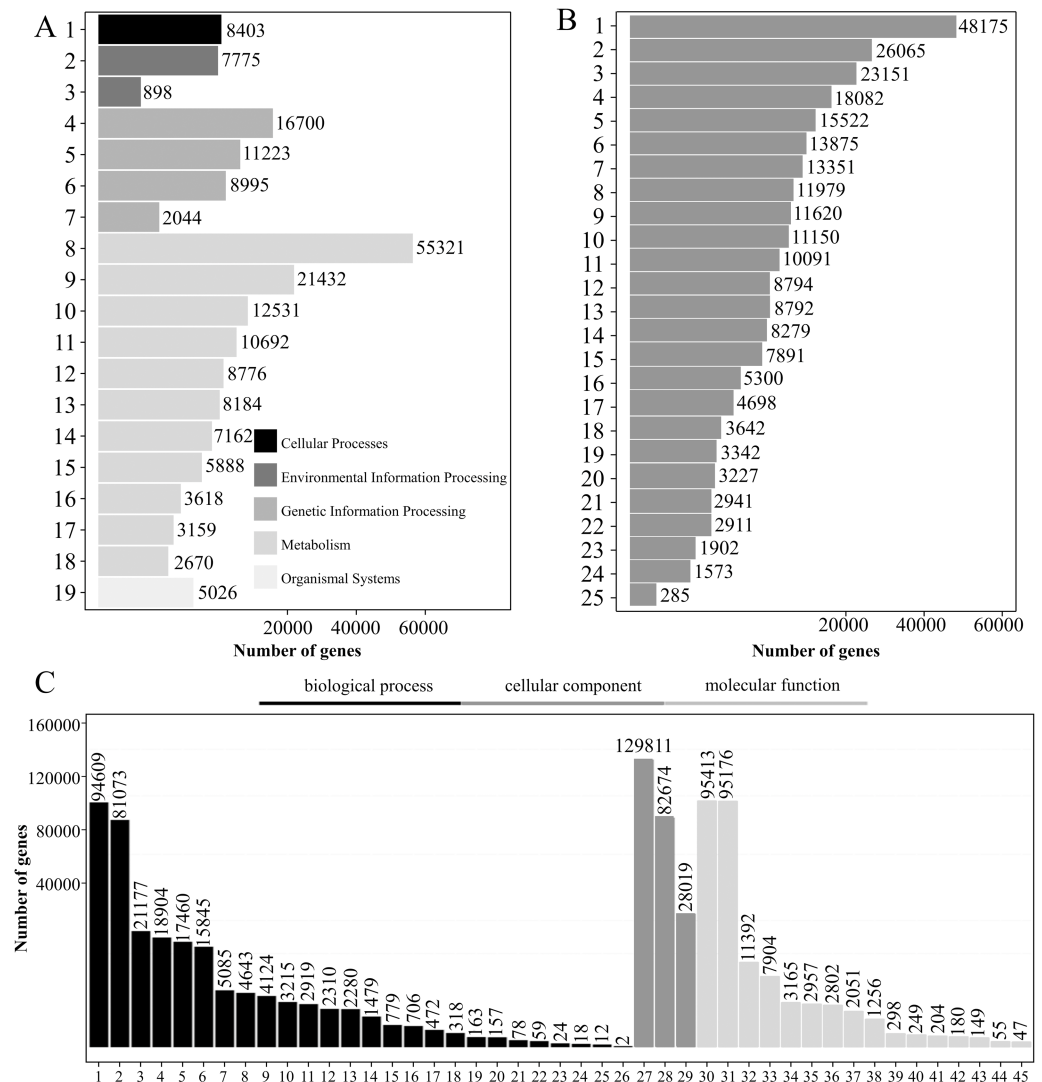


Figure 4 KEGG pathway analysis, KOG functional classification, and GO analysis of transcripts. (A–C) KEGG pathway classification (A), KOG functional classification (B), and GO functional classification (C) of transcripts. Note A: 1: Transport and catabolism; 2: Signal transduction; 3: Membrane transport; 4: Translation; 5: Folding, sorting and degradation; 6: Transcription; 7: Replication and repair; 8: Global and overview maps; 9: Carbohydrate metabolism; 10: Energy metabolism; 11: Amino acid metabolism; 12: Lipid metabolism; 13: Biosynthesis of other secondary metabolites; 14: Metabolism of cofactors and vitamins; 15: Metabolism of other amino acids; 16: Metabolism of terpenoids and polyketides; 17: Nucleotide metabolism; 18: Glycan biosynthesis and metabolism; 19: Environmental adaptation. Note B: 1: General function prediction only; 2: Signal transduction mechanisms; 3: Posttranslational modification, protein turnover, chaperones; 4: Function unknown; 5: Carbohydrate transport and metabolism; 6: Energy production and conversion; 7: Transcription; 8: Translation, ribosomal structure and biogenesis; 9: Cell wall/membrane/envelope biogenesis; 10: RNA processing and modification; 11: Lipid transport and metabolism; (continued on next page...)

Full-size DOI: [10.7717/peerj.13079/fig-4](https://doi.org/10.7717/peerj.13079/fig-4)

Figure 4 (...continued)

12: Cytoskeleton; 13: Intracellular trafficking, secretion, and vesicular transport; 14: Amino acid transport and metabolism; 15: Secondary metabolites biosynthesis, transport and catabolism; 16: Inorganic ion transport and metabolism; 17: Defense mechanisms; 18: Cell cycle control, cell division, chromosome partitioning; 19: Chromatin structure and dynamics; 20: Replication, recombination and repair; 21: Extracellular structures; 22: Coenzyme transport and metabolism; 23: Nucleotide transport and metabolism; 24: Nuclear structure; 25: Cell motility. Note C: 1: Cellular process; 2: Metabolic process; 3: Biological regulation; 4: Regulation of biological process; 5: Response to stimulus; 6: Localization; 7: Signaling; 8: Developmental process; 9: Multicellular organismal process; 10: Positive regulation of biological process; 11: Negative regulation of biological process; 12: Reproduction; 13: Reproductive process; 14: Interspecies interaction between organisms; 15: Growth; 16: Multi-organism process; 17: Immune system process; 18: Detoxification; 19: Rhythmic process; 20: Carbon utilization; 21: Locomotion; 22: Nitrogen utilization; 23: Biological adhesion; 24: Intraspecies interaction between organisms; 25: Pigmentation; 26: Carbohydrate utilization; 27: Cellular anatomical entity; 28: Intracellular; 29: Protein-containing complex; 30: Binding; 31: Catalytic activity; 32: Transporter activity; 33: Structural molecule activity; 34: Transcription regulator activity; 35: Molecular function regulator; 36: Translation regulator activity; 37: Antioxidant activity; 38: Molecular transducer activity; 39: Protein folding chaperone; 40: Small molecule sensor activity; 41: Nutrient reservoir activity; 42: Protein tag; 43: Molecular carrier activity; 44: Cargo receptor activity; 45: Toxin activity.

Table 6 Predicted results of transcription factors, transcriptional regulators, and protein kinases.

Item	Total transcripts	Transcription factors	Transcription regulators	Protein kinases
Number	354,228	5,749	2,214	4,950
Percentage		1.62%	0.63%	1.39%

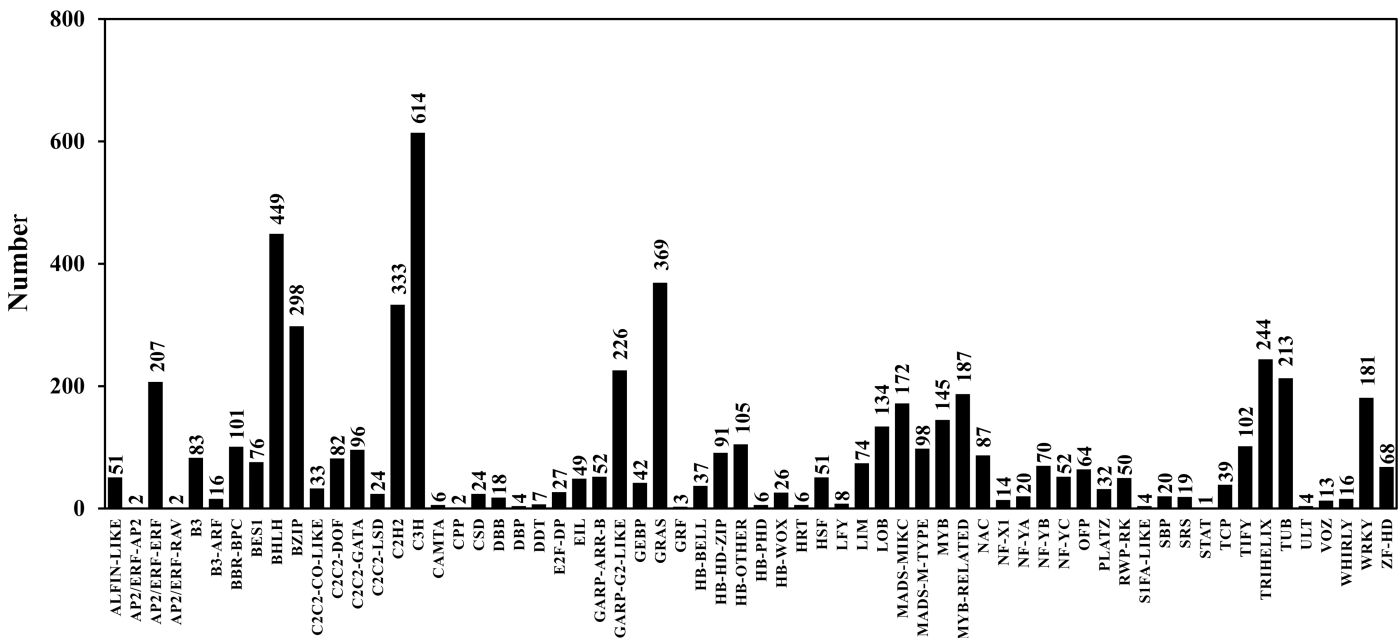


Figure 5 Transcription factor classification.

Full-size DOI: 10.7717/peerj.13079/fig-5

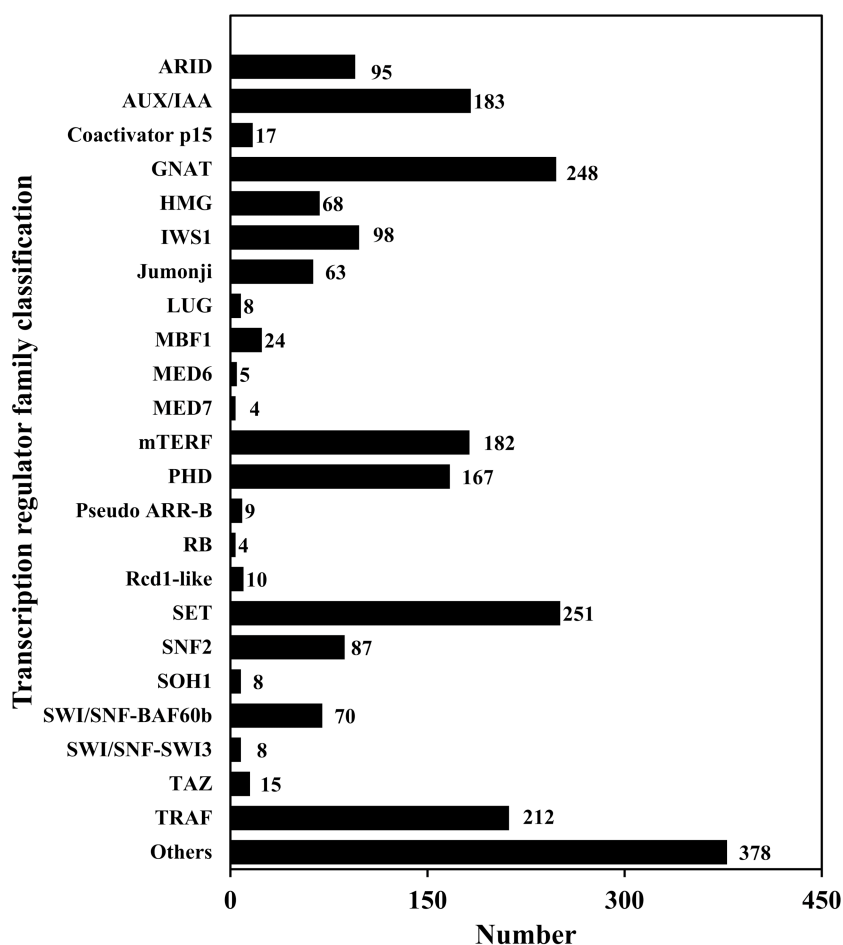


Figure 6 Transcription regulator classification.

[Full-size](#) [DOI: 10.7717/peerj.13079/fig-6](https://doi.org/10.7717/peerj.13079/fig-6)

lncRNAs prediction

106,931, 118,406, 99,672 and 115,438 lncRNAs were predicted using CPC, txCdsPredict, CNCI tools and the pfam database, respectively (Table S10). Among them, 85,487 were predicted using the 4 methods together, and 2,105, 7,652, 1,839 and 14,710 were predicted by three of the four methods alone (Fig. 8). Finally, 111,793 lncRNAs were obtained, accounting for 31.56% and 91.24% of all transcripts and non-CDSs, respectively.

Quantitative real-time polymerase chain reaction

Chlorophyll is an important pigment in plant photosynthesis, and its synthesis enzyme gene expressions are regulated by light (Lee et al., 2005; Masuda et al., 2003; Matsumoto et al., 2004; Okamoto et al., 2020; Tzvetkova-Chevolleau et al., 2007). In order to verify the reliability of the full-length transcript sequences and the annotation information, we selected nine chlorophyll synthase genes according to KEGG annotation for qRT-PCR, including *HemA*, *HemL*, *HemB*, *CHLH*, *CHLD*, *CRD*, *DVR*, *POR* and *CAO*. The results showed that the expressions of the above nine genes increased at first and then decreased with the increase of illumination (Table S11; Fig. 9). It can be seen that the chlorophyll

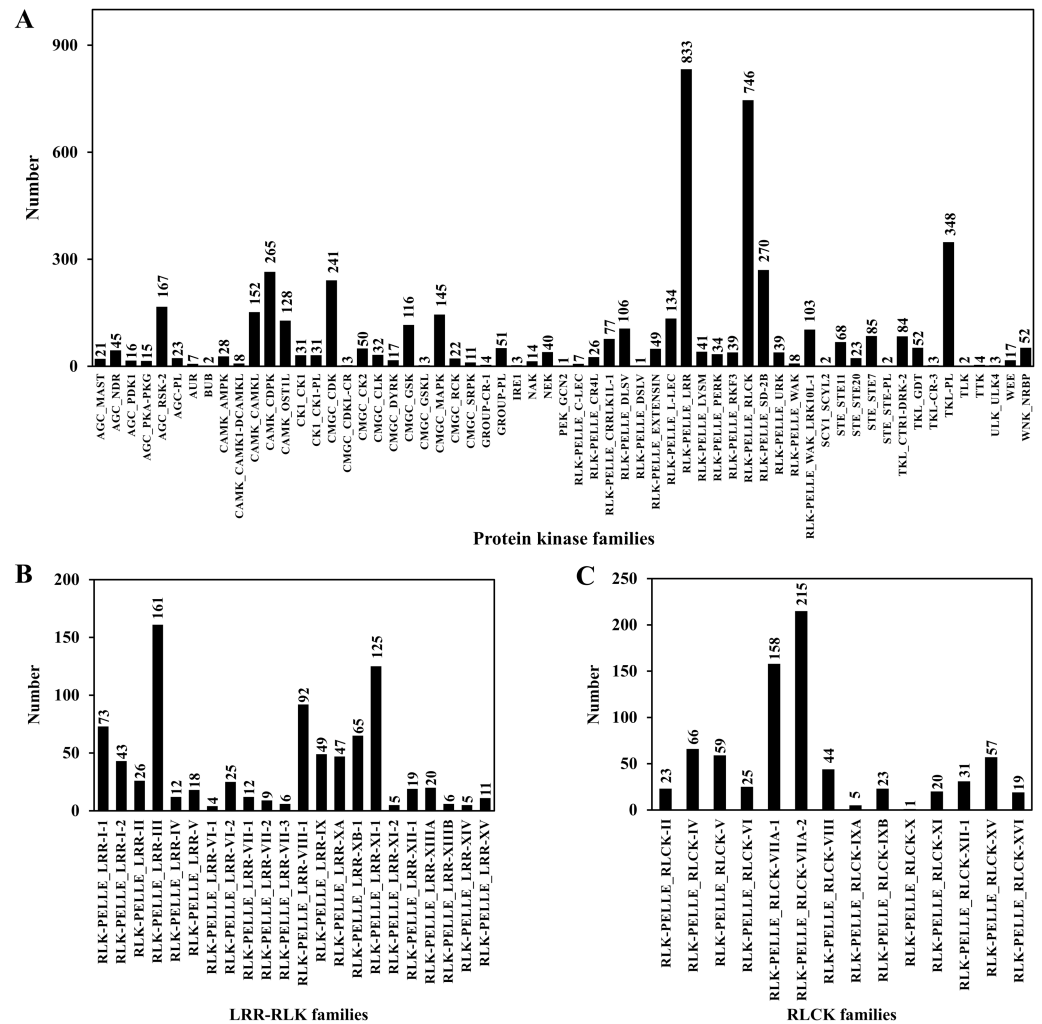


Figure 7 Protein kinase classification. (A) Protein kinase classification; (B) LRR-RLK classification; (C) RLCK classification.

Full-size [DOI: 10.7717/peerj.13079/fig-7](https://doi.org/10.7717/peerj.13079/fig-7)

synthase genes in gametophyte of *A. flabellulatum* are also regulated by light. This also shows that the data of the full-length transcripts has strong reliability.

DISCUSSION

Based on the Illumina HiSeq sequencing platform, the transcriptomes of roots and shoots of *Athyrium yokoscense* (Ukai et al., 2020), young leaves of *Asplenium nidus*, young leaves of *Asplenium komarovii* (Zhang et al., 2019) and sexual reproduction and apogamy of *Adiantum reniforme* gametophytes (Fu & Chen, 2019) were studied, and 35,681, 89,741, 77,912, 333,352, 264,791 transcripts were obtained, respectively. The sequencing read length of the Illumina HiSeq is short, the mRNAs need to be broken into short fragments before sequencing, and the complete transcripts can be obtained only by splicing. The genomes of ferns are huge, and the genes are highly redundant, so errors inevitably occur

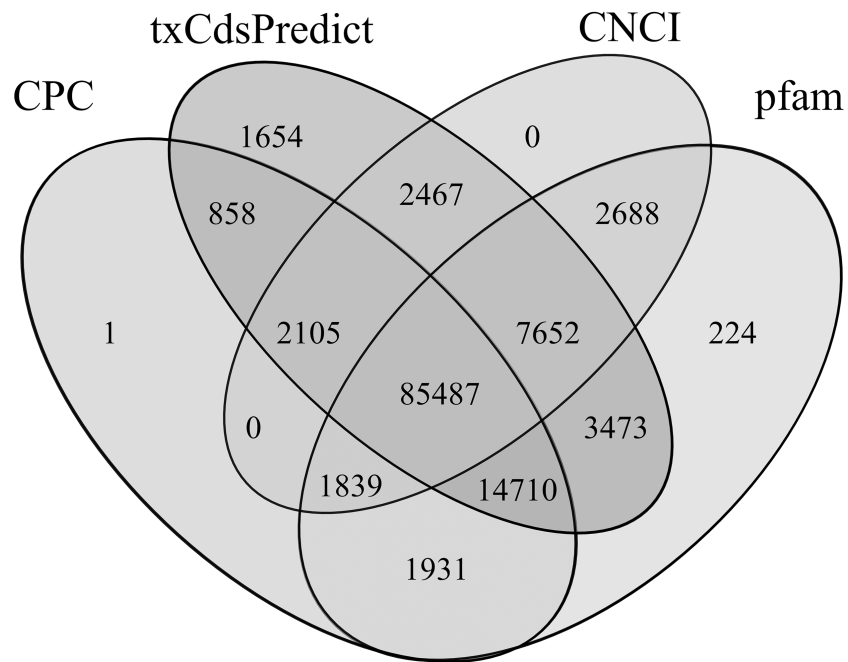


Figure 8 Venn diagram of the lncRNAs numbers predicted by CPC, txCdsPredict, and CNCI software and the Pfam database. Numbers in the overlapping areas represent the number of shared lncRNAs.

Full-size [DOI: 10.7717/peerj.13079/fig-8](https://doi.org/10.7717/peerj.13079/fig-8)

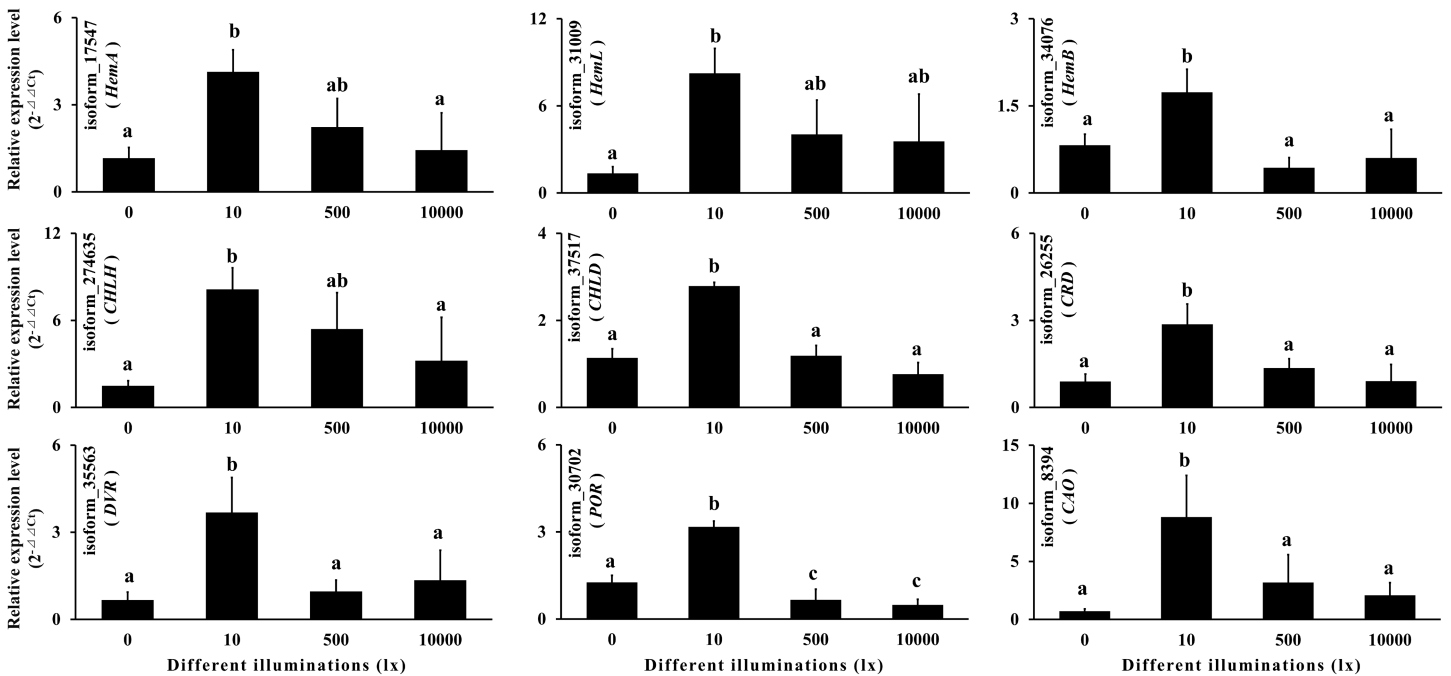


Figure 9 Quantitative real-time polymerase chain reaction. According to the KEGG annotation, nine chlorophyll synthase genes were selected for qRT-PCR. The results showed that the expressions of these nine genes first increased and then decreased with increasing illumination.

Full-size [DOI: 10.7717/peerj.13079/fig-9](https://doi.org/10.7717/peerj.13079/fig-9)

in the splicing process. Therefore, in this study, full-length transcriptome sequencing of *A. flabellulatum* gametocytes was performed using PacBio SMRT technology. This technology has very long average read lengths and can read the gene sequences completely. We believe that this technology is more suitable for species with high gene redundancy, which may be one of the reasons for the large number of transcripts obtained in this study.

SSRs, also known as short tandem repeats (STRs) or microsatellites (Xu, Sun & Li, 2010), play an important role in genetic development and gene expression regulation (Lawson & Zhang, 2006). Existing studies have shown that AG/CT is the dominant motif of di-nucleotide repeats in ferns, dicotyledons and monocotyledons (Jia et al., 2014; Jia et al., 2016; Kumpatla & Mukhopadhyay, 2005; Liu et al., 2016; Morgante, Hanafey & Powell, 2002; Zheng, Zhang & Wu, 2011), which is consistent with the SSR detection results of *A. flabellulatum* in this study. AT/AT is the second most common di-nucleotide repeat motif in dicotyledons and monocotyledons in other studies (Kumpatla & Mukhopadhyay, 2005; Zheng, Zhang & Wu, 2011), but in this experiment, AT/AT was the least common di-nucleotide repeat motif. CCG/CGG is the largest tri-nucleotide repeat motif in monocotyledons (Morgante, Hanafey & Powell, 2002; Zheng, Zhang & Wu, 2011), but is rare in dicotyledons and ferns, and the number of CCG/CGG in this study only accounted for 0.19% of the total SSRs.

The NR annotation results showed that the dominant species were *S. jiangnanensis* (10.98%, lycophytes/ Lycopodiopsida), *M. polymorpha* (11.56%, bryophytes), *P. patens* (9.45%, bryophytes), and *P. sitchensis* (8.70%, gymnosperms). *A. flabellulatum* belongs to Polypodiopsida (true ferns), which confirms that there is a certain relationship between Polypodiopsida and Lycopodiopsida in plant evolution (The Pteridophyte Phylogeny Group, 2016). In the history of plant evolution, ferns have a relatively close relationship with bryophytes and gymnosperms (Wickett et al., 2014), which is also why the transcripts can be annotated to bryophytes and gymnosperms. The NR database annotation results are consistent with the plant genetic relationship, so this study can provide similar reference gene sequences for other ferns.

TFs, TRs and PKs are important components of cell signal transduction, and also important elements in the gene regulatory network. They are widely present in different organisms and participate in various life activities (Zheng et al., 2016). Among about 25,000 genes in the angiosperm model plant *A. thaliana*, the numbers of TFs and PKs were 2,391 and 1,028, accounting for about 9.56% and 4.11%, respectively (Aglawe et al., 2012). There are about 30,000 genes in *Sorghum bicolor*, including 2,654 TFs, accounting for about 8.85% of the plant's genes (Baillo et al., 2019). Liaquat et al. (2021) detected 163,834 transcripts in *Schima superba*, and obtained 9,423 TFs, accounting for 5.75% of the transcripts. Of the 181,141 transcripts in *Huperzia serrata*, 3,391 (1.87%) were TFs (Yang et al., 2017). There were 101,448 transcripts of *Monachosorum maximowiczii*, including 1,130 TFs, accounting for 1.11% of the transcripts (Liu et al., 2016). In this study, a total of 5,749 TFs (1.62%) and 4,950 PKs (1.39%) were predicted from 354,228 full-length transcripts. It can be seen that the proportions of TFs and PKs in the gametophyte transcripts of *A. flabellulatum* are less than those in angiosperms, which may be related to the general characteristics of ferns or the relatively simple structure of fern gametophytes.

CONCLUSIONS

In this study, full-length transcriptome sequencing of *A. flabellulatum* gametocytes was performed using PacBio SMRT technology. The genomes of ferns are huge, and the genes are highly redundant, so this was a new attempt at obtaining full-length transcripts of ferns using this technology. Moreover, we predicted lncRNAs of ferns for the first time, and finally obtained 111 793 lncRNAs, accounting for 31.56% of all transcripts. This indicates that a large number of lncRNAs exist in ferns. A total of 354,228 transcripts were obtained in this study which can provide a reference gene set for subsequent gene expression quantification.

ACKNOWLEDGEMENTS

We sincerely thank reviewers and editors of the PeerJ for providing valuable suggestions on this paper.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Hainan Provincial Natural Science Foundation of China (No.319MS017), the National Natural Science Foundation of China (No.31660229), the Hainan University Scientific Research Startup Fund Project (No.kyqd1620), and the National innovation and entrepreneurship training program for college students of Hainan University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript

Grant Disclosures

The following grant information was disclosed by the authors:

Hainan Provincial Natural Science Foundation of China: No.319MS017.

National Natural Science Foundation of China: No.31660229.

Hainan University Scientific Research Startup Fund Project: No.kyqd1620.

National innovation and entrepreneurship training program for college students of Hainan University.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Zeping Cai and Jiajia Luo conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Zhenyu Xie performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Luyao Huang analyzed the data, authored or reviewed drafts of the paper, and approved the final draft.

- Zixuan Wang and Min Pan analyzed the data, prepared figures and/or tables, and approved the final draft.
- Xudong Yu and Shitao Xu conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The full-length transcript data is available in the SRA database: [PRJNA733457](https://www.ncbi.nlm.nih.gov/sra/PRJNA733457).

Data Availability

The following information was supplied regarding data availability:

The data is available at NCBI SRA: [PRJNA774117](https://www.ncbi.nlm.nih.gov/sra/PRJNA774117), [PRJNA733457](https://www.ncbi.nlm.nih.gov/sra/PRJNA733457), [PRJNA777740](https://www.ncbi.nlm.nih.gov/sra/PRJNA777740).

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.13079#supplemental-information>.

REFERENCES

- Aglawe SB, Fakrudin B, Patole CB, Bhairappanavar SB, Koti RV, Krishnaraj PU. 2012.** Quantitative RT-PCR analysis of 20 transcription factor genes of MADS, ARF, HAP2, MBF and HB families in moisture stressed shoot and root tissues of sorghum. *Physiology and Molecular Biology of Plants* **18(4)**:287–300 DOI [10.1007/s12298-012-0135-5](https://doi.org/10.1007/s12298-012-0135-5).
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990.** Basic local alignment search tool. *Journal of Molecular Biology* **215(3)**:403–410 DOI [10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Ardito F, Giuliani M, Perrone D, Troiano G, Muzio LLo. 2017.** The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *International Journal of Molecular Medicine* **40(2)**:271–280 DOI [10.3892/ijmm.2017.3036](https://doi.org/10.3892/ijmm.2017.3036).
- Baillo EH, Kimotho RN, Zhang Z, Xu P. 2019.** Transcription factors associated with abiotic and biotic stress tolerance and their potential for crops improvement. *Gene* **10(10)**:771 DOI [10.3390/genes10100771](https://doi.org/10.3390/genes10100771).
- Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. 2007.** UniProtKB/Swiss-Prot. *Methods in Molecular Biology* **406**:89–112 DOI [10.1007/978-1-59745-535-0_4](https://doi.org/10.1007/978-1-59745-535-0_4).
- Buchfink B, Xie C, Huson DH. 2015.** Fast and sensitive protein alignment using DIAMOND. *Nature Methods* **12(1)**:59–60 DOI [10.1038/nmeth.3176](https://doi.org/10.1038/nmeth.3176).
- Cai ZP, Huang Z, Wang ZX, Tao Y, Wu FH, Yu XD, Luo JJ. 2021.** Identification of the related genes on the asymmetric root growth of *Oryza sativa* induced by ethylene through transcriptome sequencing, GO and KEGG analysis. *Acta Physiologiae Plantarum* **43**:99 DOI [10.1007/s11738-021-03271-9](https://doi.org/10.1007/s11738-021-03271-9).
- Cao JG, Huang WJ, Wang QX. 2010.** Microstructural observation on the development of gametophytes and oogenesis in the fern *Adiantum flabellulatum* L. *Acta Botanica Boreali-Occidentalia Sinica* **30(4)**:702–707 DOI [10.3724/SP.J.1231.2010.06705](https://doi.org/10.3724/SP.J.1231.2010.06705).

- Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013.** Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods* **10**(6):563–569 DOI [10.1038/nmeth.2474](https://doi.org/10.1038/nmeth.2474).
- Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. 2005.** Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**(18):3674–3676 DOI [10.1093/bioinformatics/bti610](https://doi.org/10.1093/bioinformatics/bti610).
- Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, Salazar GA, Tate J, Bateman A. 2016.** The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research* **44**(D1):D279–D285 DOI [10.1093/nar/gkv1344](https://doi.org/10.1093/nar/gkv1344).
- Fu Q, Chen LQ. 2019.** Comparative transcriptome analysis of two reproductive modes in *Adiantum reniforme* var. *sinense* targeted to explore possible mechanism of apogamy. *BMC Genetics* **20**(1):55 DOI [10.1186/s12863-019-0762-8](https://doi.org/10.1186/s12863-019-0762-8).
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. 2013.** De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8**(8):1494–1512 DOI [10.1038/nprot](https://doi.org/10.1038/nprot).
- Hidalgo O, Pellicer J, Christenhusz M, Schneider H, Leitch AR, Leitch IJ. 2017a.** Is there an upper limit to genome size? *Trends in Plant Science* **22**(7):567–573 DOI [10.1016/j.tplants.2017.04.005](https://doi.org/10.1016/j.tplants.2017.04.005).
- Hidalgo O, Pellicer J, Christenhusz MJM, Schneider H, Leitch IJ. 2017b.** Genomic gigantism in the whisk-fern family (Psilotaceae): *Tmesipteris obliqua* challenges record holder *Paris japonica*. *Botanical Journal of the Linnean Society* **183**(4):509–514 DOI [10.1093/botlinnean/box003](https://doi.org/10.1093/botlinnean/box003).
- Hong F, Mo SH, Lin XY, Niu JZ, Yin J, Wei D. 2020.** The PacBio full-length transcriptome of the tea aphid as a reference resource. *Frontiers in Genetics* **11**:558394 DOI [10.3389/fgene.2020.558394](https://doi.org/10.3389/fgene.2020.558394).
- Jia XP, Deng YM, Sun XB, Liang LJ, Su JL. 2016.** De novo assembly of the transcriptome of *Neottopteris nidus* using Illumina paired-end sequencing and development of EST-SSR markers. *Molecular Breeding* **36**(7):94 DOI [10.1007/s11032-016-0519-2](https://doi.org/10.1007/s11032-016-0519-2).
- Jia XP, Sun XB, Deng YM, Liang LJ, Ye XQ. 2014.** Sequencing and analysis of the transcriptome of *Asplenium nidus*. *Acta Horticulturae Sinica* **41**(11):2329–2341 DOI [10.16420/j.issn.0513-353x.2014.11.021](https://doi.org/10.16420/j.issn.0513-353x.2014.11.021).
- Jia XP, Tang L, Mei XY, Liu HZ, Luo HR, Deng YM, Su JL. 2020.** Single-molecule long-read sequencing of the full-length transcriptome of *Rhododendron lapponicum* L. *Scientific Reports-Uk* **10**(1):6755 DOI [10.1038/s41598-020-63814-x](https://doi.org/10.1038/s41598-020-63814-x).
- Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G. 2007.** CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research* **35**:W345–W349 DOI [10.1093/nar/gkm391](https://doi.org/10.1093/nar/gkm391).

- Kumpatla SP, Mukhopadhyay S. 2005.** Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome* **48**(6):985–998 DOI [10.1139/g05-060](https://doi.org/10.1139/g05-060).
- Kuo LY, Li FW. 2019.** A roadmap for fern genome sequencing. *American Fern Journal* **109**(3):212–223 DOI [10.1640/0002-8444-109.3.212](https://doi.org/10.1640/0002-8444-109.3.212).
- Lawson MJ, Zhang L. 2006.** Distinct patterns of SSR distribution in the *Arabidopsis thaliana* and rice genomes. *Genome Biology* **7**(2):R14 DOI [10.1186/gb-2006-7-2-r14](https://doi.org/10.1186/gb-2006-7-2-r14).
- Lee S, Kim JH, Yoo ES, Lee CH, Hirochika H, An G. 2005.** Differential regulation of chlorophyll a oxygenase genes in rice. *Plant Mol Biol* **57**(6):805–818 DOI [10.1007/s11103-005-2066-9](https://doi.org/10.1007/s11103-005-2066-9).
- Li FW, Brouwer P, Carretero-Paulet L, Cheng SF, De Vries J, Delaux PM, Eily A, Koppers N, Kuo LY, Li Z, Simenc M, Small I, Wafula E, Angarita S, Barker MS, Brautigam A, De Pamphilis C, Gould S, Hosmani PS, Huang YM, Huettel B, Kato Y, Liu X, Maere S, McDowell R, Mueller LA, Nierop KGJ, Rensing SA, Robison T, Rothfels CJ, Sigel EM, Song Y, Timilsena PR, Van de Peer Y, Wang HL, Wilhelmsson PKI, Wolf PG, Xu X, Der JP, Schlupepmann H, Wong GKS, Pryer KM. 2018.** Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants* **4**(7):460–472 DOI [10.1038/s41477-018-0188-8](https://doi.org/10.1038/s41477-018-0188-8).
- Li W, Godzik A. 2006.** Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**(13):1658–1659 DOI [10.1093/bioinformatics/btl158](https://doi.org/10.1093/bioinformatics/btl158).
- Liaquat F, Munis MFH, Arif S, Haroon U, Shi J, Saqib S, Zaman W, Che S, Liu Q. 2021.** PacBio single-molecule long-read sequencing reveals genes tolerating manganese stress in *Schima superba* saplings. *Frontiers in Genetics* **12**:635043 DOI [10.3389/fgene.2021.635043](https://doi.org/10.3389/fgene.2021.635043).
- Liu L, Shu JP, Wei HJ, Zhang R, Shen H, Yan YH. 2016.** De novo transcriptome analysis of the rare fern *Monachosorum maximowiczii* (Dennstaedtiaceae) endemic to East Asia. *Biodiversity Science* **24**(12):1325–1334 DOI [10.17520/biods.2016231](https://doi.org/10.17520/biods.2016231).
- Marchant DB, Sessa EB, Wolf PG, Heo K, Barbazuk WB, Soltis PS, Soltis DE. 2019.** The C-Fern (*Ceratopteris richardii*) genome: insights into plant genome evolution with the first partial homosporous fern genome assembly. *Scientific Reports-Uk* **9**(1):18181 DOI [10.1038/s41598-019-53968-8](https://doi.org/10.1038/s41598-019-53968-8).
- Masuda T, Fusada N, Oosawa N, Takamatsu K, Yamamoto YY, Ohto M, Nakamura K, Goto K, Shibata D, Shirano Y, Hayashi H, Kato T, Tabata S, Shimada H, Ohta H, Takamiya K. 2003.** Functional analysis of isoforms of NADPH: protochlorophyllide oxidoreductase (POR), PORB and PORC, in *Arabidopsis thaliana*. *Plant Cell Physiology* **44**(10):963–974 DOI [10.1093/pcp/pcg128](https://doi.org/10.1093/pcp/pcg128).
- Matsumoto F, Obayashi T, Sasaki-Sekimoto Y, Ohta H, Takamiya K, Masuda T. 2004.** Gene expression profiling of the tetrapyrrole metabolic pathway in *Arabidopsis* with a mini-array system. *Plant Physiology* **135**(4):2379–2391 DOI [10.1104/pp.104.042408](https://doi.org/10.1104/pp.104.042408).
- Morgante M, Hanafey M, Powell W. 2002.** Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics* **30**:194–200 DOI [10.1038/ng822](https://doi.org/10.1038/ng822).

- Okamoto H, Ducreux LJM, Allwood JW, Hedley PE, Wright A, Gururajan V, Terry MJ, Taylor MA. 2020. Light regulation of chlorophyll and glycoalkaloid biosynthesis during tuber greening of potato *S. tuberosum*. *Frontiers in Plant Science* 11:753 DOI 10.3389/fpls.2020.00753.
- Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* 29:45 DOI 10.1093/nar/29.9.e45.
- Schneider H, Schuettpelz E, Pryer KM, Cranfill R, Magallón S, Lupia R. 2004. Ferns diversified in the shadow of angiosperms. *Nature* 428(6982):553–557 DOI 10.1038/nature02361.
- Sessa EB, Banks JA, Barker MS, Der JP, Duffy AM, Graham SW, Hasebe M, Langdale J, Li FW, Marchant DB, Pryer KM, Rothfels CJ, Roux SJ, Salmi ML, Sigel EM, Soltis DE, Soltis PS, Stevenson DW, Wolf PG. 2014. Between two fern genomes. *Gigascience* 3:15 DOI 10.1186/2047-217x-3-15.
- Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, Liu Y, Chen R, Zhao Y. 2013. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Research* 41(17):e166 DOI 10.1093/nar/gkt646.
- The Pteridophyte Phylogeny Group. 2016. A community-derived classification for extant lycophytes and ferns. *Journal of Systematics and Evolution* 54(6):563–603 DOI 10.1111/jse.12229.
- Thiel T, Michalek W, Varshney RK, Graner A. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical & Applied Genetics* 106(3):411–422 DOI 10.1007/s00122-002-1031-0.
- Tzvetkova-Chevolleau T, Franck F, Alawady AE, Dall'Osto L, Carrière F, Bassi R, Grimm B, Nussaume L, Havaux M. 2007. The light stress-induced protein ELIP2 is a regulator of chlorophyll synthesis in *Arabidopsis thaliana*. *The Plant Journal* 50(5):795–809 DOI 10.1111/j.1365-313X.2007.03090.x.
- Ukai Y, Inoue K, Kamada M, Teramura H, Yanagisawa S, Kitazaki K, Shoji K, Goto F, Mochida K, Yoshihara T, Shimada H. 2020. De novo transcriptome analysis reveals an unperturbed transcriptome under high cadmium conditions in the Cd-hypertolerant fern *Athyrium yokoscense*. *Genes & Genetic Systems* 95(2):65–74 DOI 10.1266/ggs.19-00056.
- Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3-new capabilities and interfaces. *Nucleic Acids Research* 40(15):e115 DOI 10.1093/nar/gks596.
- Wang JM, Yu XD, Cai ZP, Xu ST, Luo JJ, Yang ZW. 2018. Effects of brassinosteroids on gametophyte growth of *Adiantum flabellulatum* L. *Chinese Journal of Tropical Crops* 39(9):1739–1744 DOI 10.3969/j.issn.1000-2561.2018.09.009.
- Wickett NJ, Mirarab S, Nguyen N, Warnow T, Carpenter E, Matasci N, Ayyampalayam S, Barker MS, Burleigh JG, Gitzendanner MA, Ruhfel BR, Wafula E, Der JP, Graham SW, Mathews S, Melkonian M, Soltis DE, Soltis PS, Miles NW, Rothfels CJ, Pokorny L, Shaw AJ, De Gironimo L, Stevenson DW, Surek B, Villarreal JC, Roure B, Philippe H, De Pamphilis CW, Chen T, Deyholos MK, Baucom RS,

- Kutchan TM, Augustin MM, Wang J, Zhang Y, Tian Z, Yan Z, Wu X, Sun X, Wong GK, Leebens-Mack J. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences of the United States of America* **111**(45):e4859–4868 DOI [10.1073/pnas.1323926111](https://doi.org/10.1073/pnas.1323926111).
- Wu Q, Zang F, Xie X, Ma Y, Zheng Y, Zang D. 2020. Full-length transcriptome sequencing analysis and development of EST-SSR markers for the endangered species *Populus wulianensis*. *Scientific Reports* **10**(1):16249 DOI [10.1038/s41598-020-73289-5](https://doi.org/10.1038/s41598-020-73289-5).
- Xu M, Sun YG, Li HG. 2010. EST-SSRs development and paternity analysis for *Liriodendron* spp. *New Forests* **40**(3):361–382 DOI [10.1007/s11056-010-9205-0](https://doi.org/10.1007/s11056-010-9205-0).
- Yang M, You W, Wu S, Fan Z, Xu B, Zhu M, Li X, Xiao Y. 2017. Global transcriptome analysis of *Huperzia serrata* and identification of critical genes involved in the biosynthesis of huperzine A. *BMC Genomics* **18**(1):245 DOI [10.1186/s12864-017-3615-8](https://doi.org/10.1186/s12864-017-3615-8).
- Yu H, Liu M, Yin M, Shan T, Peng H, Wang J, Chang X, Peng D, Zha L, Gui S. 2021. Transcriptome analysis identifies putative genes involved in triterpenoid biosynthesis in *Platycodon grandiflorus*. *Planta* **254**(2):34 DOI [10.1007/s00425-021-03677-2](https://doi.org/10.1007/s00425-021-03677-2).
- Zhang J, Liu L, Shu JP, Jin DM, Shen H, Chen HF, Zhang R, Yan YH. 2019. Transcriptomic evidence of adaptive evolution of the epiphytic fern *Asplenium nidus*. *International Journal of Plant Genomics* **2019**:1429316 DOI [10.1155/2019/1429316](https://doi.org/10.1155/2019/1429316).
- Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, Zhang P, Banf M, Dai X, Martin GB, Giovannoni JJ, Zhao PX, Rhee SY, Fei Z. 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 DOI [10.1016/j.molp.2016.09.014](https://doi.org/10.1016/j.molp.2016.09.014).
- Zheng Y, Zhang G, Wu W. 2011. Characterization and comparison of microsatellites in gramineae. *Genomics and Applied Biology* **30**(5):513–520 DOI [10.3969/gab.030.000513](https://doi.org/10.3969/gab.030.000513).
- Zweig AS, Karolchik D, Kuhn RM, Haussler D, Kent WJ. 2008. UCSC genome browser tutorial. *Genomics* **92**(2):75–84 DOI [10.1016/j.ygeno.2008.02.003](https://doi.org/10.1016/j.ygeno.2008.02.003).