

## Article

# PPDP: A Data Portal of *Paris polyphylla* for Polyphyllin Biosynthesis and Germplasm Resource Exploration

Qixuan Su <sup>1,2</sup>, Xuan Zhang <sup>3</sup>, Jing Li <sup>2</sup>, Wenjing Yang <sup>2,4</sup>, Qiang Ren <sup>4</sup>, Xiaoyang Gao <sup>2,\*</sup> and Changning Liu <sup>1,2,5,6,\*</sup>

<sup>1</sup> School of Life Sciences, University of Science and Technology of China, Hefei 230026, China

<sup>2</sup> CAS Key Laboratory of Tropical Plant Resources and Sustainable Use, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Kunming 650223, China

<sup>3</sup> Center for Molecular Diagnosis and Precision Medicine, The Department of Clinical Laboratory, The First Affiliated Hospital of Nanchang University, Nanchang 330006, China

<sup>4</sup> College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China

<sup>5</sup> Center of Economic Botany, Core Botanical Gardens, Chinese Academy of Sciences, Xishuangbanna 666303, China

<sup>6</sup> The Innovative Academy of Seed Design, Chinese Academy of Sciences, Kunming 650223, China

\* Correspondence: gaoxiaoyang@xtbg.ac.cn (X.G.); liuchangning@xtbg.ac.cn (C.L.)

**Abstract:** *Paris polyphylla* Smith is a perennial medicinal herb with records from around 2000 years ago. Polyphyllins are the main bioactive compounds of this herb, which are found to have remarkable effects on bacteriostatic, antiphlogistic, sedative, and antitumor. However, the market demand for *P. polyphylla* is sharply increasing, and the wild resources are threatened by plundering exploitation. Integrating molecular data of *P. polyphylla* can benefit the sustainable resource exploitation. Here, we constructed PPDP (*Paris polyphylla* Data Portal) to provide a data platform for polyphyllin biosynthesis and germplasm resource research. PPDP integrates related molecular data resources, functional genomics analysis, and morphological identification. The database provides abundant data (transcriptome, CDS, lncRNA, alternative splicing, gene family, SSR, and chloroplast genome) and practical analytical tools (network construction, heatmap of expression profiles, enrichment, and pathway search) with a user-friendly interface. So far, PPDP is the first biomolecular database for the genus *Paris* plants. In the future, we will gradually add genomic data and other necessary molecular biological information to improve the database.

**Keywords:** *Paris polyphylla*; polyphyllin; germplasm resource; database



**Citation:** Su, Q.; Zhang, X.; Li, J.; Yang, W.; Ren, Q.; Gao, X.; Liu, C. PPDP: A Data Portal of *Paris polyphylla* for Polyphyllin Biosynthesis and Germplasm Resource Exploration. *Diversity* **2022**, *14*, 1057. <https://doi.org/10.3390/d14121057>

Academic Editor: Mario A. Pagnotta

Received: 1 November 2022

Accepted: 29 November 2022

Published: 1 December 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

*Paris polyphylla* Smith (Melanthiaceae) has a medicinal history of over 2000 years, it is distributed across Assam, Bangladesh, North-Central China, South-Central China, East Himalaya, Myanmar, Nepal, Qinghai, Tibet, and West Himalaya. *Paris* plants mostly grow in evergreen broad-leaved forests, bamboo forests, shrubs, or grass slopes. Currently, more than 10 varieties have been found in *P. polyphylla* [1,2]. *Paris polyphylla* var. *yunnanensis* (Franch.) Hand.-Mazz. is one of the most significant varieties of *P. polyphylla* recorded in Pharmacopoeia of the People's Republic of China [3]. In a recent monograph published in 2021, it is recognized as a species named *Paris yunnanensis* Franch. [4]. Pharmacological studies have shown that *P. polyphylla* has significant effects on hemostasis, analgesia, bacteriostasis, anti-inflammatory, and tumor cell inhibition [5]. To date, more than 210 compounds have been isolated and identified from *P. polyphylla*, including steroidal saponins, C-21 steroids, fatty acid esters, triterpenes, flavonoids,  $\beta$ -ecdysone, and polysaccharides [6,7]. Steroidal saponins, i.e., polyphyllins, are the most important active ingredients, accounting for approximately 57% of the total number of identified ingredients. Therefore, rhizoma paridis has become the key drug material of more than 80 kinds of patented medicines, such as gongxuening, yunnan baiyao, reduqing, and

jidesheng sheyao tablets [8]. However, the long growth cycle of 5–7 years, and the post-embryonic maturation of seeds result in the slow reproduction of *P. polyphylla*. Moreover, sharply increasing market demand and large-scale over-excavation driven by high medical value cause resource shortage, and also endanger the wild resources of *P. polyphylla* and other *Paris* species [9]. Therefore, understanding polyphyllin biosynthesis mechanism is significant for the effective utilization of this medicinal herb.

With the development of high-throughput sequencing, transcriptome sequencing has been widely used to mine functional genes involved in plant secondary metabolism processes, especially the specialized metabolite biosynthesis of non-model plants lacking genome sequences. The exploration of polyphyllin biosynthesis genes (PBGs) has remarkably advanced through transcriptome sequencing. Several candidate genes related to polyphyllin biosynthesis were predicted based on the transcriptome data from rhizomes [10] or leaves (stems) [11] of *P. polyphylla*. Meanwhile, another 25 transcripts encoding 17 key enzymes related to polyphyllin biosynthesis were identified from the transcriptome of *P. polyphylla* tissue mixtures [12]. More specifically, a gene encoding cytochrome P450 monooxygenase (P450) was identified, which catalyzes the oxidative 5,6-spiroketalization of cholesterol to produce diosgenin [13]. A total of 137 PBGs, 74 transcription factor genes, and 1 transporter gene associated with polyphyllin biosynthesis and accumulation were identified in our previous study [14]. Additionally, polyphyllin biosynthesis is found to be in response to tissue-specific combinatorial development cues through RNA-Seq analysis [14,15]. Nonetheless, molecular mechanisms underlying polyphyllin biosynthesis and accumulation are still unclear.

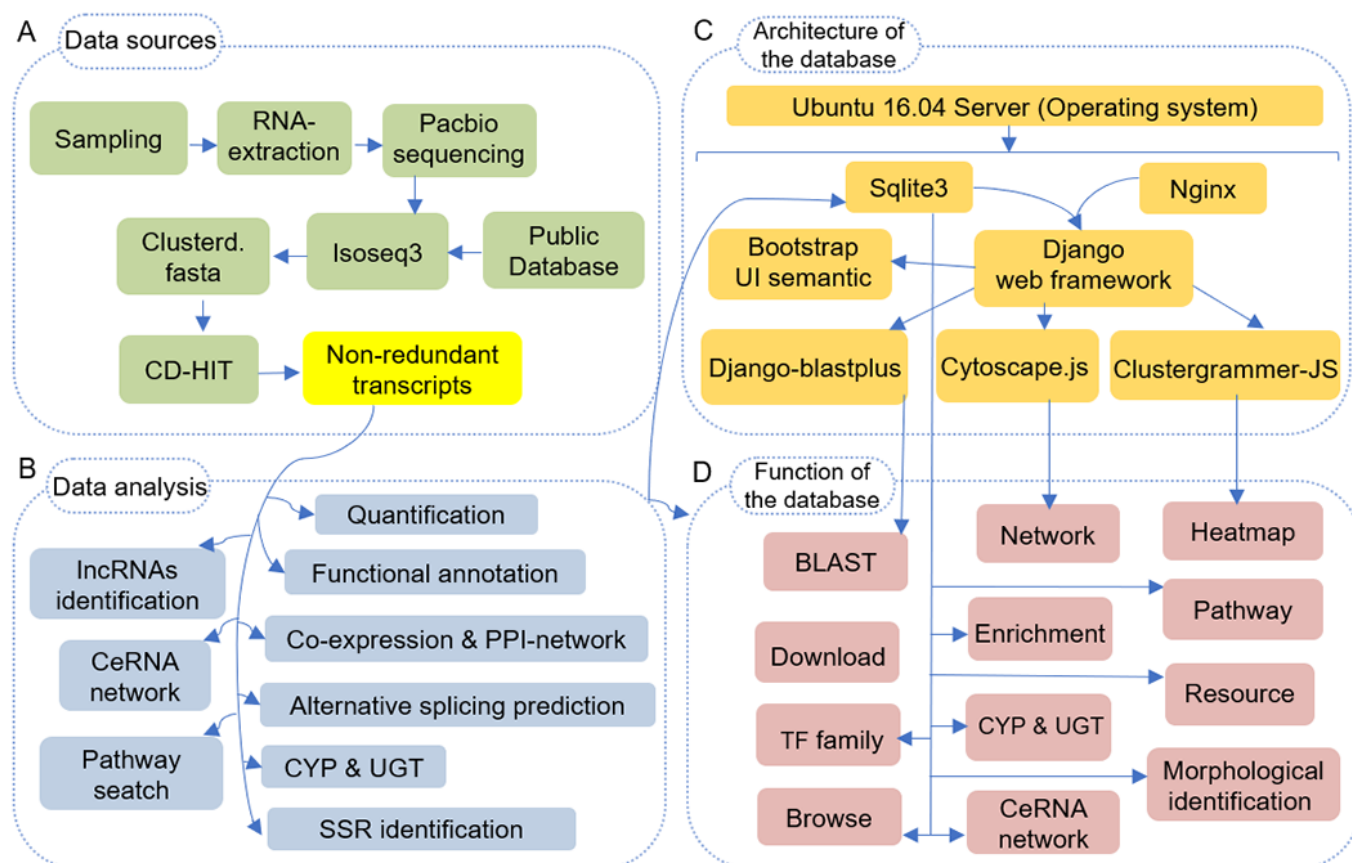
A database that integrates omics data and analytical tools is very beneficial to unravel specialized metabolite biosynthesis mechanism. Some medicinal plant databases such as the Ginseng Genome Database [16] facilitate terpenoid biosynthesis research and molecular breeding [17,18]. However, there is no database available for *P. polyphylla* so far. Genetic information of *P. polyphylla* is primarily from next-generation sequencing (NGS) transcriptome data [19,20] and very few third-generation sequencing (TGS) transcriptome data [21]. In this study, we constructed the first data portal PPDP (<http://ppdp.liu-lab.com>) for *P. polyphylla*, mainly based on the PacBio SMRT-based RNA-Seq and Illumina-based RNA-Seq data. The current version of the database provides diverse data, including 36,504 genes, 9020 lncRNAs, 7029 alternative splicing (AS) events, 25,767 SSRs and 49 valid SSR markers, 108 chloroplast genomes, 931 genes from 50 transcription factor (TF) families, 93 CYP and 56 UGT gene candidates, and gene expression profiles at different growth stages. A total of 186 PBGs were identified, of which 88 genes contained SSRs. PPDP provides a kit of practical analytical tools, such as network construction, heatmap drawing, pathway search, etc., with a user-friendly interface. In all, PPDP provides new molecular data and helps to gain further insight into polyphyllin biosynthesis.

## 2. Materials and Methods

### 2.1. Sampling and PacBio SMRT Sequencing

Four fresh tissue samples were collected from six 7-year-old plants of *P. polyphylla* var. *yunnanensis*, which were planted under the same field management in the greenhouse of Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences (Kunming). Each fresh tissue sample was a mixture, which combined the same tissue collected from three plants with similar size and status to obtain credible sequencing data. The *P. polyphylla* var. *yunnanensis* plants in this study were obtained through seed germination and seedling cultivation for years. The seeds were kindly provided by Yunnan Yuxin Agriculture and Forestry Biological Technology Co., Ltd (Yunnan, China). Seven-year-old is an important stage in the agriculture cultivation of *Paris* plants. The rhizomes of *P. polyphylla* plants from seed propagation are often harvested at the seventh year. This age is chosen for providing a representative dataset and a reasonable reference in metabolism biosynthesis research. The species identification was confirmed by Professor Li Heng (Kunming Institute of Botany, CAS). RNA from roots and stems were pooled into one sample for SMRT sequencing.

mRNA was purified using magnetic beads with Oligo (dT), and the library was prepared according to the Isoform Sequencing protocol (Iso-Seq) using the Clontech SMARTer PCR cDNA Synthesis Kit and the BluePippin Size Selection System protocol as described by Pacific Biosciences (PN 100-092-800-03). The sequencing was performed using the PacBio Sequel II platform after confirming library quality. Flowers and leaves were sequenced separately in the same way after RNA extraction. An additional PacBio isoform sequencing dataset of *P. polyphylla* var. *yunnanensis* was downloaded from The National Genomics Data Center (NGDC) (accession number: CRA004081) [21] (Figure 1A).



**Figure 1.** Implementation of PPDP. (A) RNA-Seq of sampled tissues and data pre-processing. (B) The main data analysis process. (C) The basic architecture of PPDP. (D) The main functions and related tools of PPDP.

## 2.2. Transcriptome Determination and Evaluation

Three sets of PacBio raw data of *P. polyphylla* var. *yunnanensis* (two sequenced in this study and one downloaded from the public database) were processed according to the Iso-Seq3 pipeline (<https://github.com/PacificBiosciences/IsoSeq> (accessed on 10 November 2021)). First, circular consensus sequence (CCS) was generated from subread BAM files with parameter settings: min\_length 50, max\_length 15,000, min-rq 0.8. Then, CCS reads were defined as full-length (FL) and non-full-length (NFL) isoforms, depending on the presence or absence of 5' primers, 3' primers, and poly(A) tails. Primers, poly(A) tails, and rapid concatemers were also removed. Finally, full-length non-chimeric (FLNC) reads were clustered to generate high-quality (HQ) isoforms and low-quality (LQ) isoforms. In addition, the three FLNC reads were merged by clustering in the expectation of a new transcript of higher quality. De-redundancy using CD-HIT v4.8.1 [22] and the removal of transcripts shorter than 200 bp yielded the final non-redundant transcripts, respectively. Together with our previous NGS transcripts based on Illumina sequencing data from 4 developmental stages and 5 tissues in *P. polyphylla* var. *yunnanensis* [13,14], all five transcripts were assessed using BUSCO v5.0.0 [23].

### 2.3. Functional Annotation and Classification

Functional annotation was performed to gain an insight into the biological context of the transcripts. All the non-redundant transcripts were searched against Nr and KEGG using BLAST v2.2.25 [24] with an E-value threshold of  $1 \times 10^{-5}$ . The annotation of GO, Swiss-Prot, and Pfam were determined using Trinotate v3.2.2 (<http://trinotate.github.io/> (accessed on 20 November 2021)). Transcripts were also annotated by eggno-mapper v2.6.1 [25]. Coding sequences (CDS) and open reading frames (ORFs) were identified with TransDecoder v5.5.0 (<https://github.com/TransDecoder/TransDecoder/releases> (accessed on 20 November 2021)).

### 2.4. SSR Analysis

SSR determination was performed using MISA v2.1 [26]. Six types of SSRs including mono-, di-, tri-, tetra-, penta-, and hexa-nucleotide repeats with minimum repeat numbers of 10, 6, 5, 5, 5, and 5 were identified from the transcriptome. The distance between adjacent SSRs < 100 bp was defined as compound SSR. Further, functional annotations of transcripts containing SSR were obtained. In addition, valid SSR markers and chloroplast genomes were collected from the literature and NCBI database, respectively.

### 2.5. Identification of AS and lncRNAs

The transcripts gtf file raw.gtf was first obtained. Then, the AS events were predicted using SUPPA v2.3 with default parameters [27]. lncRNAs are non-coding transcripts longer than 200 bp that regulate gene expression in various forms. lncRNAs in PPDP were identified by evaluating the coding potential of transcripts using CNCI v2 [28], CPC2 v 0.1 [29], PLEK v1.2 [30], and Pfam. Finally, transcripts at any two intersections of the prediction results of CNCI, CPC2, PLEK, and Pfam were used as candidate lncRNAs.

### 2.6. Prediction and Construction of ceRNA Network

In plants, miRNA genes are a class of highly conserved gene families. As no whole genome of *P. polyphylla* is available, *Arabidopsis thaliana* mature miRNAs from the PmiREN database [31] were used to predict the ceRNA network. The psRNATarget web server was used to predict targets [32]. The ceRNA score was calculated with our python script to determine whether lncRNA–mRNA pairs were ceRNAs ( $p$ -value < 0.05) [33]. The intersection was taken to obtain the final predicted ceRNA regulatory relationship pair and the ceRNA network was acquired. Then the ceRNA network was constructed using Cytoscape v3.9.1 [34].

### 2.7. Explore the Biosynthesis Pathway of Secondary Metabolites in *P. polyphylla*

The following biosynthesis pathways, including “fatty acid elongation”, “steroid biosynthesis”, “starch and sucrose metabolism”, “terpenoid backbone biosynthesis”, “phenylpropanoid biosynthesis”, and “flavonoid biosynthesis” in *P. polyphylla* were explored by defining *A. thaliana* as a template. Protein sequences of *A. thaliana* were obtained from TAIR [35] and gene accession numbers of the above-mentioned pathways were obtained from the KEGG PATHWAY Database. Then, the biosynthetic genes from the pathways of *P. polyphylla* were acquired by BLAST with the *A. thaliana* sequences and data filtration with an E-value threshold of  $1 \times 10^{-5}$ .

### 2.8. Prediction of Transcription Factor and CYP&UGT

The TF families were identified by mapping the protein sequences to the database PlantTFDB 5.0 [36]. As CYPs and UGTs play very important roles in secondary metabolism, they are also critical in polyphyllin biosynthesis. A total of 288 CYP protein sequences from cytochrome P450 (<http://drnelson.uthsc.edu/CytochromeP450.html> (accessed on 21 March 2022)); and 122 UGT protein sequences from *Arabidopsis thaliana* cytochromes P450, cytochromes b<sub>5</sub>, NADPH-cytochrome P450 reductases, and β-Glucosidases site (<http://www.p450.kvl.dk/UGT.shtml> (accessed on 21 March 2022)) were used to search for



homologous CYPs and UGTs in *P. polyphylla* (E-value  $< 1 \times 10^{-5}$ ). Hidden Markov models (accession numbers: PF00067 and PF00201) were obtained from the Pfam [37] based on the evolutionary conservation of gene family domains. They were used as the queries to search against the protein sequences of *P. polyphylla* using Hmmssearch v3.3.2 [38]. The protein sequences of screened complete ORFs were further submitted to NCBI-CDD and Pfam to confirm the domain. Finally, candidate CYPs and UGTs were obtained.

## 2.9. Construction of Co-Expression Network and PPI Network

Quantification of the merged transcripts was achieved by RSEM [39] using the Illumina sequencing data. The co-expression between two genes was estimated by using the traditional Pearson correlation coefficient. Gene pairs with correlation values  $> 0.8$  and adjusted  $p$  values  $< 0.01$  were considered to show co-expression [40]. The protein interactions of *A. thaliana* were collected from three databases, namely AtPID v5.0 [41], AtPIN v9.0 [42], and PAIR v3.0 [43] and the literature [44–46]. A total of 18,037 *A. thaliana* genes and 241,468 interactions were obtained. Orthologous groups between the *P. polyphylla* and *A. thaliana* were detected using InParanoid v4.2 [47] with default parameters. The PPI network in *P. polyphylla* was inferred from the *A. thaliana* PPI network by homology mapping (Figure 1B).

## 2.10. Development of the Morphological Identification

First of all, the description of morphological characteristics of different species of the genus *Paris*, various specimen information of *Paris* in the CHV China Digital Herbarium (<https://primulaworld.blogspot.com/2015/12/the-chinese-virtual-herbarium-cvh.html> (accessed on 2 May 2020)), expert opinions, and the morphological classification feature data of *Paris* were collected. Then, the random forest classifier [48] was used to train the model of the preprocessed data after comparison and screening. Finally, the trained model was encapsulated and saved for use [49].

## 2.11. Implementation of PPDP

PPDP was developed using Django web framework (<http://www.djangoproject.com/> (accessed on 23 February 2021)) based on Python programming. The website runs on a Linux Server. The Nginx web server (<https://www.nginx.com/> (accessed on 10 March 2021)) and SQLite (<https://www.sqlite.org/> (accessed on 20 March 2021)) were used as the web server and database server, respectively. The web front end was developed using Bootstrap (<https://www.ubuntu.com/> (accessed on 30 March 2021)) and the Semantic UI framework (<https://semantic-ui.com/> (accessed on 15 April 2021)). JavaScript libraries, highchart.js (<https://www.highcharts.com/> (accessed on 10 May 2022)), and datatables (<https://datatables.net/> (accessed on 3 April 2021)) were applied for rendering interactive graphs and tables. The network visualization was implemented using Cytoscape.js v.3.3 [50]. The online BLAST server was built with Django-blastplus (<https://pypi.python.org/pypi/django-blastplus/> (accessed on 20 May 2021)) (Figure 1C). Finally, PPDP provides a series of user-friendly functions, such as browse, download, BLAST, heatmap, network, enrichment, pathway search, TF family, etc. (Figure 1D).

# 3. Results

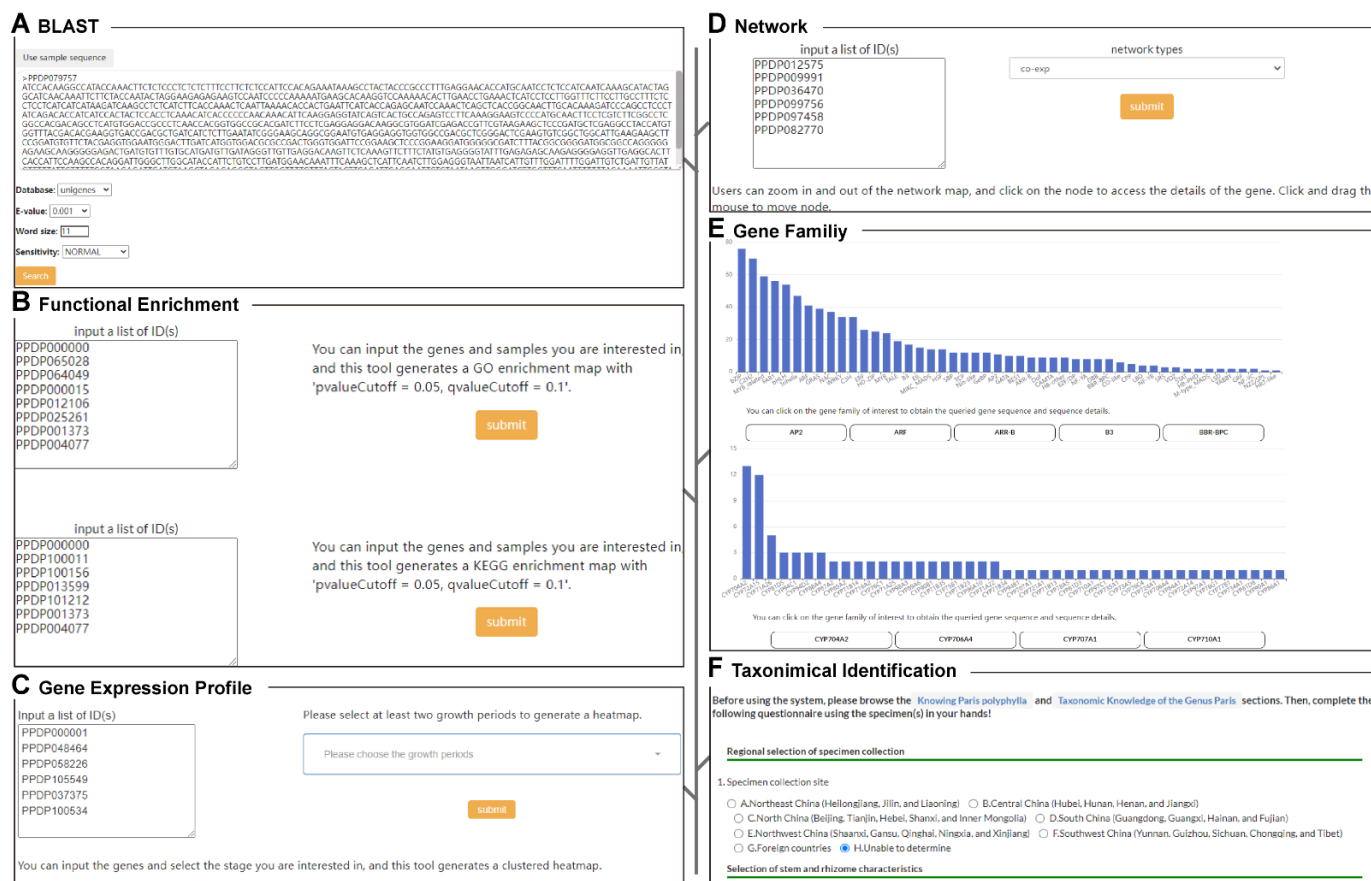
## 3.1. The Web Interface of PPDP

### 3.1.1. Browse PPDP

All the genes were listed on the browse page and the detailed information page can be accessed by clicking the gene ID. For each gene in the browse, comprehensive information was provided, including multiple gene functional annotations, AS events, heatmaps of gene expression, and co-expression sub-networks. The number of co-expressed genes displayed is optional in the gene expression section.

### 3.1.2. Visualization of Heatmap, Network, and Enrichment

The ‘Enrichment’ module provides users with GO and KEGG enrichment for genes of interest (Figure 2B). The expression patterns of genes at different growth periods are available in the ‘Heatmap’ module (Figure 2C). The ‘Network’ module enables to extract a sub-network of user-specified genes from the global co-expression network (Figure 2D). Additionally, the ‘BLAST’ is a user-friendly tool for interacting with the NCBI BLAST+ toolkits (Figure 2A).



**Figure 2.** Web interface of PPDP. (A) Finding homologous genes of *P. polyphylla* using sequence similarity search with BLAST. (B) KEGG and GO enrichment of genes of interest. (C) Viewing gene expression profiles at different growth stages. (D) Prediction of co-expression network and PPI-network. (E) TF and CYP gene family statistics and the specific gene sequences. (F) Morphological identification of *P. polyphylla* and other *Paris* species.

### 3.1.3. Pathway Search

Plant secondary metabolites play an important role in plant growth and development, biotic and abiotic stresses, and mediating interactions with other organisms [51–53]. Particularly, specialized secondary metabolites of medicinal plants are responsible for medicinal activity. A total of 61, 36, 25, 171, 36, and 128 genes involved in terpenoid backbone biosynthesis, steroid biosynthesis, flavonoid biosynthesis, starch and sucrose metabolism, fatty acid elongation, and phenylpropanoid biosynthesis in *A. thaliana* were obtained. Correspondingly, a total of 38, 21, 12, 99, 16, and 49 genes of these pathways in *P. polyphylla* were acquired through BLAST. The reference pathway map, related biosynthetic genes in *A. thaliana* and *P. polyphylla*, and their corresponding KEGG enzymes are presented by selecting a wanted pathway.

### 3.1.4. TF, CYP, and UGT Family

In the 'TF family' module, the distribution of TF families is represented as a bar chart. The TF family of interest can be clicked to obtain the queried sequences and each sequence is linked to its detail page. The 'CYP & UGT' module demonstrates the CYP and UGT gene families in the same way (Figure 2E).

### 3.1.5. Other Data Resource

An abundance of *P. polyphylla* datasets can be freely downloaded through the 'Downloads' section, including transcript sequences, CDS sequences, lncRNA sequences, and annotations. The 'Resource' section provides valid SSR markers and chloroplast genomes of *P. polyphylla* and other *Paris* species. In the 'Identification System' section, users can taxonomically identify their samples by the morphological characters of *Paris* specimens collected, such as features of roots, stems, leaves, sepals, flowers, and fruits (Figure 2F). Moreover, pictures of the taxonomic characteristics are supplied for useful reference.

## 3.2. Database Statistics and Use Case

### 3.2.1. General Properties of PacBio Sequencing Data and Evaluation of Transcriptome

A total of 110,666 HQ isoforms and 824 LQ isoforms were acquired, and 36,812 transcripts were attained after clustering and de-redundancy. A total of 36,504 high-quality non-redundant transcripts were defined in *P. polyphylla* after filtering transcripts < 200 bp. The average length of the transcriptome determined in this study was 2008 bp, with an N50 of 2330 bp and a GC content of 45.88% (Table 1).

**Table 1.** Summary of the transcriptome of *P. polyphylla* var. *yunnanensis*.

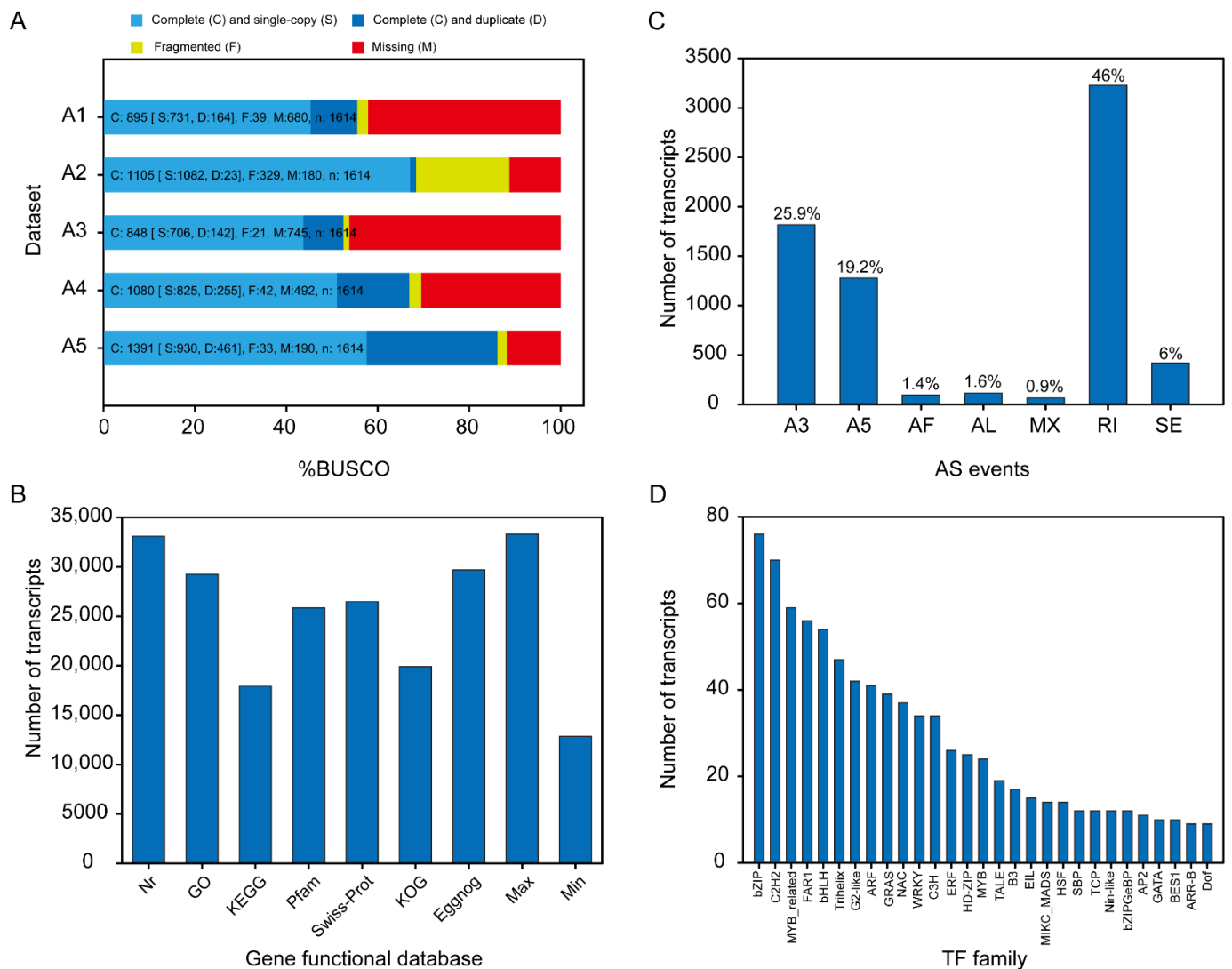
Item	Value
High-quality isoforms	110,666
Low-quality isoforms	824
Non-redundant transcripts	36,812
High-quality non-redundant transcripts	36,504
Total bases (bp)	73,313,381
N50 (bp)	2330
Average length (bp)	2008
Reads mapping (%)	86.80
Percent GC (%)	45.88

According to BUSCO assessment results, the merged transcriptome has the most complete sequences and single-copy complete sequences, followed by the Illumina-based transcriptome (Figure 3A). In addition, 82.50% of the predicted genes from the merged transcriptome indicated relatively high completeness of the merged transcriptome assembly. Furthermore, 86.8% of the Illumina reads were mapped to the newly merged transcriptome, demonstrating a high level of transcriptome coverage. Because it was of higher quality, it was used for subsequent analyses.

### 3.2.2. Data Summary

Approximately 90% of the genes were annotated to at least one of the seven public databases. Specifically, 33,100, 29,249, 17,926, 25,855, 26,493, 19,907, and 29,696 transcripts were annotated to NR, GO, KEGG, Pfam, Swiss-prot, Kog, and eggNOG, accounting for 90.67%, 80.1%, 49.1%, 70.8%, 72.6%, 54.53%, and 81.3% of the total transcripts, respectively (Figure 3B). A total of 7029 transcripts had multiple types AS. Among them, those with retained introns (RI) had the largest number (3230), accounting for 46% (Figure 3C). The result is consistent with previous reports indicating that RI is the most frequent type of AS events in plants [54,55]. Additionally, a total of 931 candidate genes from 50 TF families were found (Figure 3D). The top five TF families are bZIP, C2H2, MYB\_related,

FAR1, and bHLH. Moreover, 100,201 interactions between miRNAs, lncRNAs, and mRNAs (157 miRNAs, 879 lncRNAs, and 4922 mRNAs) were identified.

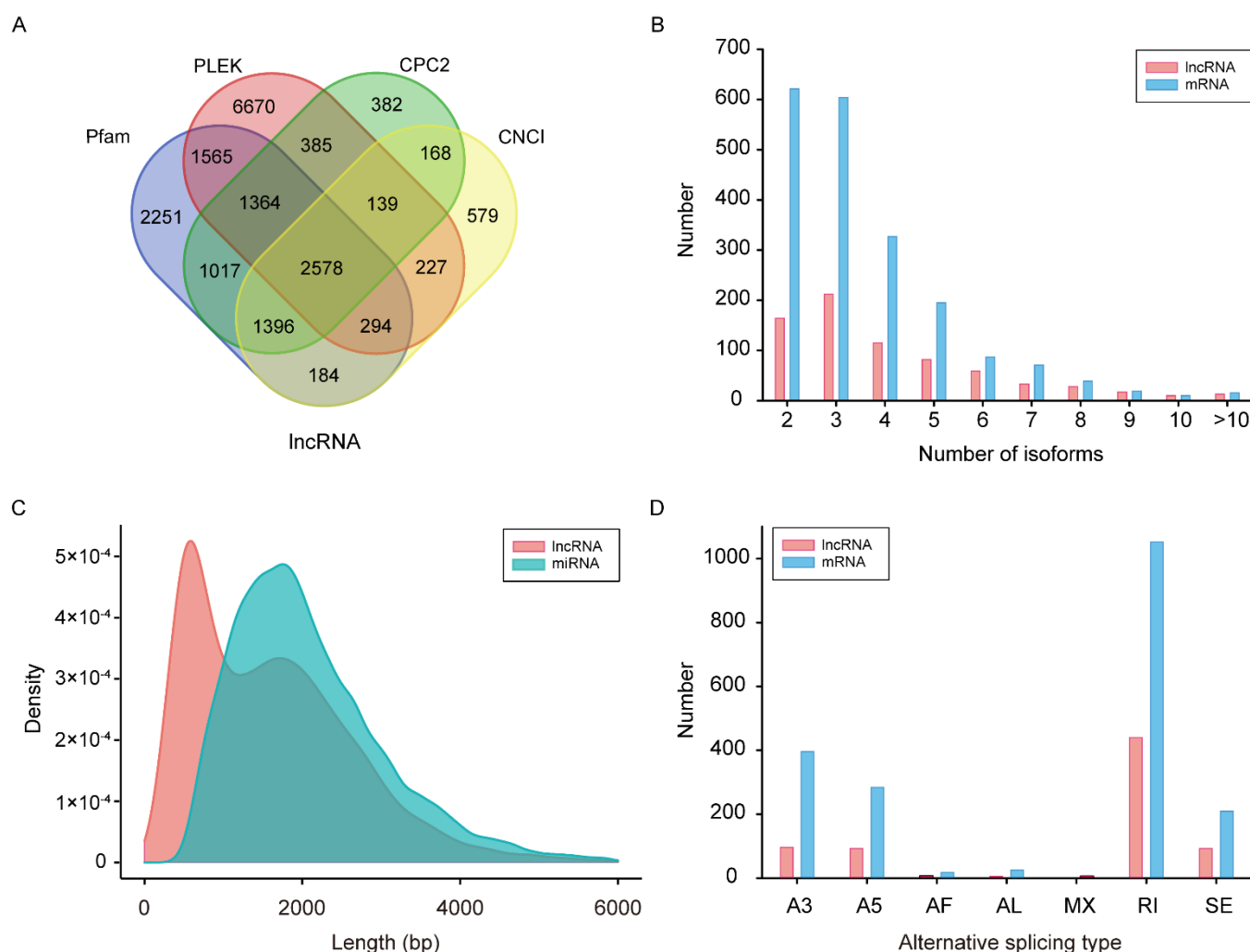


**Figure 3.** Transcriptome evaluation and annotation. **(A)** BUSCO assessment results for five transcriptomes with different datasets and assembly strategies. A1: PacBio-based RNA-Seq data from public database; A2: Illumina-based RNA-Seq data from our previous work; A3: PacBio-based RNA-Seq data from roots and stems; A4: PacBio-based RNA-Seq data from flowers and leaves; A5: merging all PacBio-based RNA-Seq data; C: complete orthologs; S: single copy orthologs; D: duplicated orthologs; F: fragmented orthologs; M: missing orthologs. **(B)** Statistics of functional annotations. **(C)** Classification of the AS events. A3: alternative 3' splice sites; A5: alternative 5' splice sites; AF: alternative first exon; AL: alternative last exon; MX: mutually exclusive exons; RI: retained introns; SE: skipping exon events. **(D)** The top 30 TF families identified in *P. polyphylla*.

### 3.2.3. Comparison of lncRNAs and mRNAs

A total of 9020 lncRNAs were identified in the *P. polyphylla* transcriptome using CPC2, CNCI, PLEK, and Pfam (Figure 4A). According to the length distribution, the average length of mRNAs (2008 bp) was longer than that of lncRNAs (1613 bp) (Figure 4C). Compared to mRNAs, lncRNAs featured fewer isoforms and AS events (Figure 4B,D).





**Figure 4.** Distribution and features of lncRNA. (A) Venn diagram of predicted lncRNAs. (B) Distribution of isoform numbers of lncRNA and mRNA. (C) Density distribution of lncRNA and mRNA length. (D) Distribution of AS events of lncRNA and mRNA.

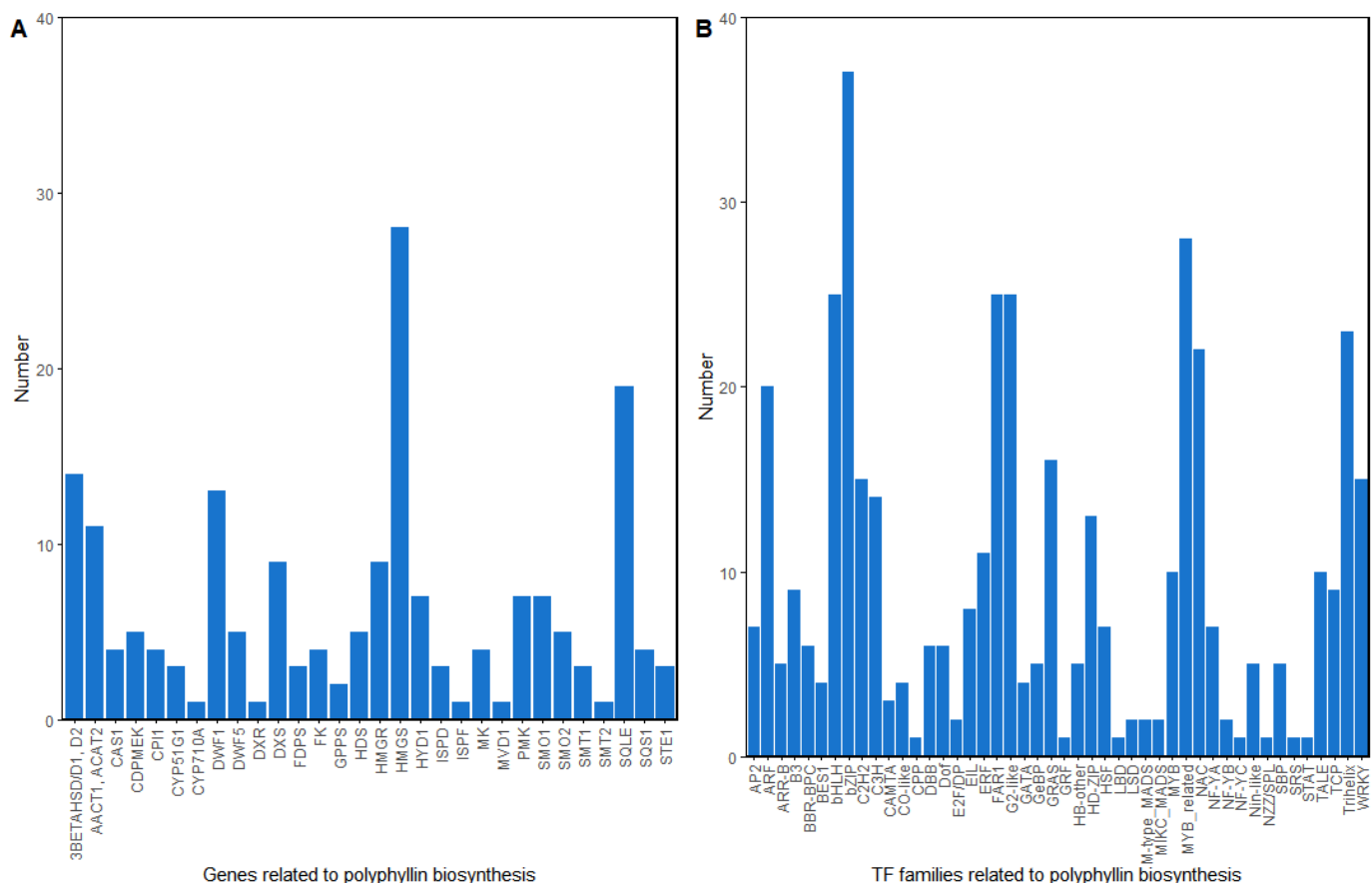
### 3.2.4. Exploration of the Polyphyllin Biosynthesis Pathway

A total of 186 candidate genes involved in the upstream of the polyphyllin biosynthesis pathway were identified (Figure 5A). These genes participate in two branch pathways of plant saponin biosynthesis: the cytosolic mevalonate (MVA) pathway and the plastidial 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway. Quantitatively, more genes were related to the MVA pathway than to the MEP pathway. The transcriptional regulation of polyphyllin biosynthesis was also predicted in this study. It showed that a total of 432 TF genes (Figure 5B) may participate in the regulation of polyphyllin biosynthesis (Supplementary Materials Figure S1).

### 3.2.5. Data Related to Germplasm Resources

A total of 25,767 potential SSRs were identified from 36,504 transcripts. Among them, dinucleotides (46.75%) were the most abundant, followed by mononucleotides (39.45%) and trinucleotides (12.14%). Moreover, 4027 compound SSRs were identified (Table 2). After functional annotation, 16,120 SSR-containing transcripts had 12,626, 7678, 11,106, 11,465, and 14,412 homologous sequences in GO, KEGG, Pfam, Swiss-prot, and NR, respectively. In total, 186 PBGs were identified from the transcriptome, of which 97 contained SSRs. Additionally, valid SSR markers and the related primers of *P. polyphylla* were collected from reference. There were 14 and 35 valid SSR markers for *P. polyphylla* var. *chinensis* [56] and

*P. polyphylla* var. *yunnanensis* [57–59], respectively. Similarly, *P. polyphylla* var. *chinensis* is also considered as a species, named *Paris chinensis* Franch. in the monograph [4]. As the chloroplast genome is an ideal system for plant phylogeny study and an efficient marker source in germplasm resource investigation, chloroplast genomes of the *Paris* species were also collected from the public database. A total of 108 chloroplast genomes of *P. polyphylla* and other *Paris* species have been added in PPDP.



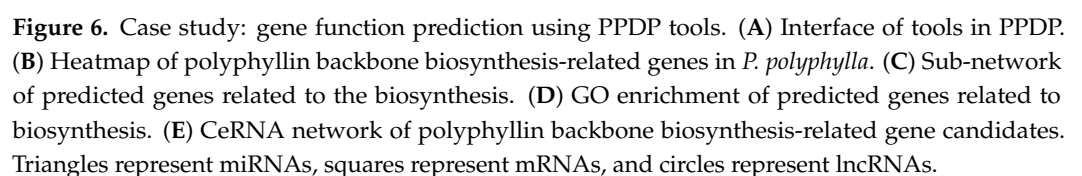
**Figure 5.** Genes and TF genes involved in polyphyllin biosynthesis. (A) Putative polyphyllin biosynthetic genes. (B) Putative TF genes related to the biosynthesis.

**Table 2.** Statistics of SSRs from the transcriptome.

Item	Value
Total number of sequences examined	36,504
Total size of examined sequences (bp)	73,313,381
Total number of identified SSRs	25,767
SSR-containing sequences	16,120
sequences containing more than 1 SSR	6169
SSRs present in compound formation	4027
Mononucleotide repeats	10,164
Dinucleotide repeats	12,045
Trinucleotide repeats	3129
Tetranucleotide repeats	88
Pentanucleotide repeats	71
Hexanucleotide repeats	270

### 3.2.6. A Case Study of Mining Genes Related to Polyphyllin Synthesis Using PPDP

PPDP is a comprehensive data platform with molecular data and analytical tools, such as BLAST, network construction, enrichment analysis, etc. Here, we demonstrate how to mine genes involved in the polyphyllin biosynthesis pathway. Polyphyllins belong to steroidal saponins, which share a similar biosynthesis backbone to that of terpenoids. From KEGG, 61 terpenoid backbone biosynthesis-related genes of *A. thaliana* were obtained, and their corresponding sequences from TAIR. Then, 38 optimal polyphyllin biosynthesis-related gene candidates in *P. polyphylla* were identified using PPDP BLAST. The expression profiles of these genes in different tissues at different growth stages can be viewed using the 'Heatmap' function (Figure 6A,B). Apparently, polyphyllin biosynthesis-related genes are expressed differentially in tissues across growth stages, especially PPDP043478, PPDP051101, and PPDP057972, which are highly expressed in leaves at pollination and fruit stages, and also in the stem at the pollination stage. A sub-network of polyphyllin backbone biosynthesis can be constructed using the 'Network Construction', indicating the existence of connections between polyphyllin biosynthesis-related genes in *P. polyphylla* (Figure 6C). Additionally, these candidate genes can be examined using 'GO Enrichment'. As expected, there was significant enrichment related to polyphyllin biosynthesis. The most abundant GO entries are isoprenoid biosynthesis processes, followed by sterol biosynthesis and terpenoid biosynthesis processes (Figure 6D). The ceRNA network of 61 genes is queried using the 'ceRNA network'. Only PPDP057217 has eight lncRNAs as its ceRNAs, indicating that these eight lncRNAs may be involved in polyphyllin backbone biosynthesis in *P. polyphylla*. (Figure 6E). Finally, the SSRs distribution of the related genes can be retrieved using the 'SSR search'. The 16 SSRs retrieved have no tetranucleotide and pentanucleotide repeats; however, they are very uniformly distributed in type mononucleotide, dinucleotides, trinucleotides, and compound repeats, which are four, five, three, and four, respectively.



Databases typically consolidate redundant-rich data, and provide many useful analytical tools [60]. Therefore, databases can greatly facilitate the research of various growth and development processes, genetic research and breeding, and secondary metabolite synthesis

pathways of plants. The *Lonicera japonica* functional genomics database (LjaFGD) includes a *Lonicera japonica* genome and 77 sets of transcriptomes, converging multiple tools for the purpose of gene functional analysis and mining of *Lonicera japonica* [61]. GinkgoDB is a comprehensive database with multi-dimensional research resources of ginkgo, which contains two versions of genomes, expression profiles, distribution information, monitoring data, and morphological photos [62]. It assists the research, development, and conservation of the entire community of ginkgo. The Citrus Pan-Genome to Breeding Database (CPBD) covers 23 genomes of 17 citrus species, 4038 sets of transcriptomes of 13 horticultural species, variations of 167 citrus resource materials, and DNA methylome of 44 citrus samples at different tissues and developmental stages [63]. Practical analysis tools are also provided in CPBD, including gene search, BLAST, gene ID conversion, KEGG/GO enrichment, CRISPR design, and genome-wide association analysis (GWAS). PPDP currently lacks phenotype–genotype association data and genomic data. It is because that GWAS/QTL studies in *P. polyphylla* have not been carried out and the large genome size of *P. polyphylla* is unavailable at present. Therefore, we need continued efforts in this area.

So far, PPDP is the first database for the *Paris* plants. Based on PacBio SMRT-based RNA-Seq and Illumina-based RNA-Seq data, PPDP contains a variety of datasets, including 36,504 genes, 9020 lncRNAs, 7029 AS events, 25,767 SSRs and 49 valid SSRs, 108 chloroplast genomes, 931 genes from 50 TF families, 93 CYP and 56 UGT gene candidates, and gene expression profiles at different growth stages. PPDP provides a range of practical analytical tools, such as BLAST online, heatmap drawing, network construction, and pathway search. In conclusion, PPDP provides new molecular data and contributes to further understanding of polyphyllin biosynthesis and germplasm resource research.

In the future, as related research increases and advances, PPDP will continue to be upgraded to collect more data and provide more functions. Our goal is that PPDP can facilitate more aspects of research on *P. polyphylla*.

## 5. Conclusions

PPDP is a user-friendly data platform that integrates functional genomic analyses, molecular data resources, and morphological identification of *P. polyphylla*. Currently, the analytical tools of PPDP v1.0 are mainly based on transcriptome data. Notably, it offers a kit of necessary tools for functional analyses of databases, including co-expression, PPI network prediction, heatmap, functional enrichment, pathway search, online BLAST, and other general tools. PPDP can contribute to functional genomic analyses and germplasm resource research on *P. polyphylla*, and it will help research of polyphyllin biosynthesis and regulation. In the future, we will add more molecular biology data to progressively improve the database.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/d14121057/s1>, Figure S1: co-expression network of polyphyllins biosynthesis genes (PBGs) and the related TF genes. The blue block indicates the PBG and the yellow block indicates TF gene; Table S1: characteristics of the transcriptomic data in this study; Table S2: information on valid SSRs collected.

**Author Contributions:** Conceptualization, C.L. and X.G.; software, Q.S., X.Z. and Q.R.; formal analysis, Q.S.; investigation, Q.S. and X.G.; resources, X.Z. and W.Y.; data curation, J.L.; writing—original draft preparation, Q.S.; writing—review and editing, X.G. and C.L.; visualization, Q.S. and X.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (Grant No. 31970609, 31800273); ‘Crop Varietal Improvement and Insect Pests Control by Nuclear Radiation’; Yunnan Fundamental Research Projects (Grant No. 202001AT070114); Startup Fund from Xishuangbanna Tropical Botanical Garden; ‘Top Talents Program in Science and Technology’ from Yunnan Province. The funders had no role in the study design, data collection, analysis and interpretation, or preparation of the manuscript. Publication costs are funded by ‘Crop Varietal Improvement and Insect Pests Control by Nuclear Radiation’.



**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available in the NCBI database with accession number PRJNA893196.

**Acknowledgments:** We are very grateful to Heng Li (Kunming Institute of Botany, Chinese Academy of Sciences) for her taxonomic consultation. We thank the following people for their kind help in this study: Xiaoke Jiang and Jiazhi Liu (Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences). We thank the Institutional Center for Shared Technologies and Facilities of Xishuangbanna Tropical Botanical Garden, CAS for providing the computer resources and technical support.

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

## References

- Li, H. *The Genus Paris Plants*; Science Press: Beijing, China, 2008; pp. 1–36.
- Chase, M.W.; Christenhusz, M.J.M.; Fay, M.F.; Byng, J.W.; Judd, W.S.; Soltis, D.E.; Mabberley, D.J.; Sennikov, A.N.; Soltis, P.S.; Stevens, P.F. An Update of the Angiosperm Phylogeny Group Classification for the Orders and Families of Flowering Plants: APG IV. *Bot. J. Linn. Soc.* **2016**, *181*, 1–20.
- Committee of National Pharmacopoeia. *Pharmacopoeia of the People's Republic of China (I)*; China Medical Science and Technology Press: Beijing, China, 2020.
- Ji, Y. *A Monograph of Paris (Melanthiaceae): 1–203*; Science Press: Beijing, China, 2021; pp. 1–203.
- Liu, Z.; Li, N.; Gao, W.; Man, S.; Yin, S.; Liu, C. Comparative study on hemostatic, cytotoxic and hemolytic activities of different species of *Paris* L. *J. Ethnopharmacol.* **2012**, *142*, 789–794. [[CrossRef](#)] [[PubMed](#)]
- Wei, J.-C.; Gao, W.-Y.; Yan, X.-D.; Wang, Y.; Jing, S.-S.; Xiao, P.-G. Chemical Constituents of Plants from the Genus *Paris*. *Chem. Biodivers.* **2014**, *11*, 1277–1297. [[CrossRef](#)] [[PubMed](#)]
- Wu, X.; Wang, L.; Wang, G.-C.; Wang, H.; Dai, Y.; Yang, X.-X.; Ye, W.-C.; Li, Y.-L. Triterpenoid saponins from rhizomes of *Paris polyphylla* var. *yunnanensis*. *Carbohydr. Res.* **2013**, *368*, 1–7. [[CrossRef](#)]
- Xin, G.; Ruoshi, L.; Duan, B.; Ying, W.; Min, F.A.N.; Shuang, W.; Zhang, H.; Xia, C. Advances in research on chemical constituents and pharmacological effects of *Paris* genus and prediction and analysis of quality markers. *Chin. Tradit. Herb. Drugs* **2019**, *50*, 4838–4852.
- Cunningham, A.; Brinckmann, J.; Bi, Y.-F.; Pei, S.-J.; Schippmann, U.; Luo, P. *Paris* in the spring: A review of the trade, conservation and opportunities in the shift from wild harvest to cultivation of *Paris polyphylla* (Trilliaceae). *J. Ethnopharmacol.* **2018**, *222*, 208–216. [[CrossRef](#)]
- Liu, T.; Li, X.; Xie, S.; Wang, L.; Yang, S. RNA-seq analysis of *Paris polyphylla* var. *yunnanensis* roots identified candidate genes for saponin synthesis. *Plant Divers.* **2016**, *38*, 163–170. [[CrossRef](#)]
- Li, B.; Peng, L.; Sun, X.; Huang, W.; Wang, N.; He, Y.; Shi, X.; Liu, Y.; Zhang, P.; Yang, X.; et al. Organ-specific transcriptome sequencing and mining of genes involved in polyphyllin biosynthesis in *Paris polyphylla*. *Ind. Crop. Prod.* **2020**, *156*, 112775. [[CrossRef](#)]
- Yang, Z.; Yang, L.; Liu, C.; Qin, X.; Liu, H.; Chen, J.; Ji, Y. Transcriptome analyses of *Paris polyphylla* var. *chinensis*, *Ypsilandra thibetica*, and *Polygonatum kingianum* characterize their steroidal saponin biosynthesis pathway. *Fitoterapia* **2019**, *135*, 52–63. [[CrossRef](#)]
- Christ, B.; Xu, C.; Xu, M.; Li, F.-S.; Wada, N.; Mitchell, A.J.; Han, X.-L.; Wen, M.-L.; Fujita, M.; Weng, J.-K. Repeated evolution of cytochrome P450-mediated spiroketal steroid biosynthesis in plants. *Nat. Commun.* **2019**, *10*, 1–11. [[CrossRef](#)]
- Gao, X.; Su, Q.; Li, J.; Yang, W.; Yao, B.; Guo, J.; Li, S.; Liu, C. RNA-Seq analysis reveals the important co-expressed genes associated with polyphyllin biosynthesis during the developmental stages of *Paris polyphylla*. *BMC Genom.* **2022**, *23*, 1–17. [[CrossRef](#)]
- Gao, X.; Zhang, X.; Chen, W.; Li, J.; Yang, W.; Zhang, X.; Li, S.; Liu, C. Transcriptome analysis of *Paris polyphylla* var. *yunnanensis* illuminates the biosynthesis and accumulation of steroidal saponins in rhizomes and leaves. *Phytochemistry* **2020**, *178*, 112460. [[CrossRef](#)] [[PubMed](#)]
- Jayakodi, M.; Choi, B.-S.; Lee, S.-C.; Kim, N.-H.; Park, J.Y.; Jang, W.; Lakshmanan, M.; Mohan, S.V.G.; Lee, D.-Y.; Yang, T.-J. Ginseng Genome Database: An open-access platform for genomics of *Panax ginseng*. *BMC Plant Biol.* **2018**, *18*, 62. [[CrossRef](#)] [[PubMed](#)]
- Fan, H.; Li, K.; Yao, F.; Sun, L.; Liu, Y. Comparative transcriptome analyses on terpenoids metabolism in field- and mountain-cultivated ginseng roots. *BMC Plant Biol.* **2019**, *19*, 82. [[CrossRef](#)]
- Tien, N.Q.D.; Ma, X.; Man, L.Q.; Chi, D.T.K.; Huy, N.X.; Nhut, D.-T.; Rombauts, S.; Ut, T.; Loc, N.H. De novo whole-genome assembly and discovery of genes involved in triterpenoid saponin biosynthesis of *Vietnamese ginseng* (*Panax vietnamensis* Ha et Grushv.). *Physiol. Mol. Biol. Plants* **2021**, *27*, 2215–2229. [[CrossRef](#)] [[PubMed](#)]
- Yin, Y.; Gao, L.; Zhang, X.; Gao, W. A cytochrome P450 monooxygenase responsible for the C-22 hydroxylation step in the *Paris polyphylla* steroidal saponin biosynthesis pathway. *Phytochemistry* **2018**, *156*, 116–123. [[CrossRef](#)]
- Qi, J.J.; Zheng, N.; Zhang, B.; Sun, P.; Hu, S.N.; Xu, W.J.; Ma, Q.; Zhao, T.Z.; Zhou, L.L.; Qin, M.J.; et al. Mining genes involved in the stratification of *Paris Polyphylla* seeds using high-throughput embryo Transcriptome sequencing. *BMC Genom.* **2013**, *14*, 358. [[CrossRef](#)]
- Hua, X.; Song, W.; Wang, K.; Yin, X.; Hao, C.; Duan, B.; Xu, Z.; Su, T.; Xue, Z. Effective prediction of biosynthetic pathway genes involved in bioactive polyphyllins in *Paris polyphylla*. *Commun. Biol.* **2022**, *5*, 50. [[CrossRef](#)]
- Li, W.; Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [[CrossRef](#)]

23. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [\[CrossRef\]](#)
24. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Cantalapiedra, C.P.; Hernández-Plaza, A.; Letunic, I.; Bork, P.; Huerta-Cepas, J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* **2021**, *38*, 5825–5829. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Beier, S.; Thiel, T.; Münch, T.; Scholz, U.; Mascher, M. MISA-web: A web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585. [\[CrossRef\]](#)
27. Trincado, J.L.; Entizne, J.C.; Hysenaj, G.; Singh, B.; Skalic, M.; Elliott, D.J.; Eyra, E. SUPPA2: Fast, accurate, and uncertainty-aware differential splicing analysis across multiple conditions. *Genome Biol.* **2018**, *19*, 40. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Sun, L.; Luo, H.; Bu, D.; Zhao, G.; Yu, K.; Zhang, C.; Liu, Y.; Chen, R.; Zhao, Y. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* **2013**, *41*, e166. [\[CrossRef\]](#)
29. Kang, Y.-J.; Yang, D.-C.; Kong, L.; Hou, M.; Meng, Y.-Q.; Wei, L.; Gao, G. CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res.* **2017**, *45*, W12–W16. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Li, A.; Zhang, J.; Zhou, Z. PLEK: A tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinform.* **2014**, *15*, 311. [\[CrossRef\]](#)
31. Guo, Z.; Kuang, Z.; Wang, Y.; Zhao, Y.; Tao, Y.; Cheng, C.; Yang, J.; Lu, X.; Hao, C.; Wang, T.; et al. PmiREN: A comprehensive encyclopedia of plant miRNAs. *Nucleic Acids Res.* **2019**, *48*, D1114–D1121. [\[CrossRef\]](#)
32. Dai, X.; Zhuang, Z.; Zhao, P.X. psRNATarget: A plant small RNA target analysis server (2017 release). *Nucleic Acids Res.* **2018**, *46*, W49–W54. [\[CrossRef\]](#)
33. Li, J.H.; Liu, S.; Zhou, H.; Qu, L.-H.; Yang, J.-H. starBase v2. 0: Decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res.* **2014**, *42*, D92–D97. [\[CrossRef\]](#)
34. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Poole, R.L. The TAIR database. *Methods Mol. Biol.* **2007**, *406*, 179–212. [\[PubMed\]](#)
36. Guo, A.-Y.; Chen, X.; Gao, G.; Zhang, H.; Zhu, Q.-H.; Liu, X.-C.; Zhong, Y.-F.; Gu, X.; He, K.; Luo, J. PlantTFDB: A comprehensive plant transcription factor database. *Nucleic Acids Res.* **2007**, *36*, D966–D969. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Finn, R.D.; Bateman, A.; Clements, J.; Coghill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; et al. Pfam: The protein families database. *Nucleic Acids Res.* **2013**, *42*, D222–D230. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Finn, R.D.; Clements, J.; Eddy, S.R. HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res.* **2011**, *39*, W29–W37. [\[CrossRef\]](#)
39. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323. [\[CrossRef\]](#)
40. Chen, W.; Li, J.; Huang, S.; Li, X.; Zhang, X.; Hu, X.; Xiang, S.; Liu, C. GCEN: An Easy-to-Use Toolkit for Gene Co-Expression Network Analysis and lncRNAs Annotation. *Curr. Issues Mol. Biol.* **2022**, *44*, 1479–1487. [\[CrossRef\]](#)
41. Li, P.; Zang, W.; Li, Y.; Xu, F.; Wang, J.; Shi, T. AtPID: The overall hierarchical functional protein interaction network interface and analytic platform for *Arabidopsis*. *Nucleic Acids Res.* **2010**, *39*, D1130–D1133. [\[CrossRef\]](#)
42. Brandão, M.M.; Dantas, L.L.; Silva-Filho, M.C. AtPIN: *Arabidopsis thaliana* Protein Interaction Network. *BMC Bioinform.* **2009**, *10*, 454. [\[CrossRef\]](#)
43. Lin, M.; Shen, X.; Chen, X. PAIR: The predicted *Arabidopsis* interactome resource. *Nucleic Acids Res.* **2010**, *39*, D1134–D1140. [\[CrossRef\]](#)
44. Jones, A.M.; Xuan, Y.; Xu, M.; Wang, R.S.; Ho, C.H.; Lalonde, S.; You, C.H.; Sardi, M.I.; Parsa, S.A.; Smith-Valle, E.; et al. Border control—A membrane-linked interactome of *Arabidopsis*. *Science* **2014**, *344*, 711–716. [\[CrossRef\]](#)
45. *Arabidopsis* Interactome Mapping Consortium. Evidence for Network Evolution in an *Arabidopsis* Interactome Map. *Science* **2011**, *333*, 601–607. [\[CrossRef\]](#)
46. Mukhtar, M.S.; Carvunis, A.-R.; Dreze, M.; Eppele, P.; Steinbrenner, J.; Moore, J.; Tasan, M.; Galli, M.; Hao, T.; Nishimura, M.T.; et al. Independently Evolved Virulence Effectors Converge onto Hubs in a Plant Immune System Network. *Science* **2011**, *333*, 596–601. [\[CrossRef\]](#) [\[PubMed\]](#)
47. O'Brien, K.P.; Remm, M.; Sonnhammer, E.L.L. Inparanoid: A comprehensive database of eukaryotic orthologs. *Nucleic Acids Res.* **2005**, *33*, D476–D480. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
49. Ren, Q.; Sang, S.Y.; Liu, C.N. Design and implementation of *Paris* plants online classification and identification system. *Computer Era* **2020**, *9*, 72–75.
50. Franz, M.; Lopes, C.T.; Huck, G.; Dong, Y.; Sumer, O.; Bader, G.D. Cytoscape.js: A graph theory library for visualisation and analysis. *Bioinformatics* **2015**, *32*, 309–311. [\[CrossRef\]](#)
51. Izhaki, I. Emodin—A secondary metabolite with multiple ecological functions in higher plants. *New Phytol.* **2002**, *155*, 205–217. [\[CrossRef\]](#)
52. Khare, S.; Singh, N.B.; Singh, A.; Hussain, I.; Niharika, K.; Yadav, V.; Bano, C.; Yadav, R.K.; Amist, N. Plant secondary metabolites synthesis and their regulations under biotic and abiotic constraints. *J. Plant Biol.* **2020**, *63*, 203–216. [\[CrossRef\]](#)

53. Bartwal, A.; Mall, R.; Lohani, P.; Guru, S.K.; Arora, S. Role of Secondary Metabolites and Brassinosteroids in Plant Defense Against Environmental Stresses. *J. Plant Growth Regul.* **2012**, *32*, 216–232. [[CrossRef](#)]
54. Zhang, G.; Guo, G.; Hu, X.; Zhang, Y.; Li, Q.; Li, R.; Zhuang, R.; Lu, Z.; He, Z.; Fang, X.; et al. Deep RNA sequencing at single base-pair resolution reveals high complexity of the rice transcriptome. *Genome Res.* **2010**, *20*, 646–654. [[CrossRef](#)] [[PubMed](#)]
55. Shen, Y.; Zhou, Z.; Wang, Z.; Li, W.; Fang, C.; Wu, M.; Ma, Y.; Liu, T.; Kong, L.-A.; Peng, D.-L.; et al. Global Dissection of Alternative Splicing in Paleopolyploid Soybean. *Plant Cell* **2014**, *26*, 996–1008. [[CrossRef](#)]
56. Wang, H.M.; Ruan, C.J.; Liang, J.H.; Han, P.; Yan, R.; Dong, X.; Liu, L.Y.; Shuai, X.C.; Liu, H.H. Development of SSR Markers of Fanjingshan *Paris polyphylla* Smith var. *chinensis* Based on High Throughput RNA-seq. *Mol. Plant Breed.* **2019**, *18*, 6059–6065.
57. Gao, X.; Su, Q.; Yao, B.; Yang, W.; Ma, W.; Yang, B.; Liu, C. Development of EST-SSR Markers Related to Polyphyllin Biosynthesis Reveals Genetic Diversity and Population Structure in *Paris polyphylla*. *Diversity* **2022**, *14*, 589. [[CrossRef](#)]
58. Chen, Z.S.Z.; Tian, B.; Cai, C.T. Genetic diversity of *Paris polyphylla* var. *yunnanensis* by SSR marker. *Chin. Tradit. Herb. Drugs* **2017**, *9*, 1834–1838.
59. Yang, W.Z.; Xu, Z.L.; Yang, S.B.; Yang, M.Q.; Zuo, Y.M.; Yang, T.M.; Zhang, J.Y. Transferability Analysis of EST-SSR Marker of Three Plants to *Paris polyphylla* Smith var. *yunnanensis* (Franch.) Hand Mazz. *Southwest China J. Agric. Sci.* **2014**, *4*, 1686–1690.
60. Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P. From data mining to knowledge discovery in databases. *AI Mag.* **1996**, *17*, 37.
61. Xiao, Q.; Li, Z.; Qu, M.; Xu, W.; Su, Z.; Yang, J. LjaFGD: *Lonicera japonica* functional genomics database. *J. Integr. Plant Biol.* **2021**, *63*, 1422–1436. [[CrossRef](#)]
62. Gu, K.J.; Lin, C.F.; Wu, J.J.; Zhao, Y.P. GinkgoDB: An ecological genome database for the living fossil, *Ginkgo biloba*. *Database J. Biol. Databases Curation* **2022**, *2022*, baac046. [[CrossRef](#)]
63. Liu, H.; Wang, X.; Liu, S.; Huang, Y.; Guo, Y.-X.; Xie, W.-Z.; Liu, H.; Qamar, M.T.U.; Xu, Q.; Chen, L.-L. Citrus Pan-genome to Breeding Database (CPBD): A comprehensive genome database for citrus breeding. *Mol. Plant* **2022**, *15*, 1503–1505. [[CrossRef](#)]