Medicinal
Plant Biology

ARTICLE

Open Access

https://doi.org/10.48130/mpb-0024-0012
Medicinal Plant Biology **2024**, 3: e013

A chromosome-level *Pinellia ternata* genome assembly provides insight into the evolutionary origin of ephedrine and acrid raphide formation

Tao Xue[1#], Haifeng Jia[2,3#], Meng Wang[2,3#], Yanting Zhang[2,3], Xiao Liu[1], Qiujie Chao[1], Fenglan Zhao[1], Zhuang Meng[4], Jianping Xue[1*], Jishan Lin[5*] and Yongbo Duan[1*]

[1] *Anhui Provincial Engineering Laboratory for Efficient Utilization of Featured Resource Plants, College of Life Sciences, Huaibei Normal University, Huaibei 235000, Anhui, China*
[2] *College of Agriculture, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China*
[3] *Center for Genomics and Biotechnology, Fujian Agriculture and Forestry University, Fuzhou 350002, Fujian, China*
[4] *College of Life Science, Shihezi University, Shihezi 832003, Xinjiang, China*
[5] *Key Laboratory of Biology and Genetic Resources of Tropical Crops, Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Haikou 571101, Hainan, China*
[#] Authors contributed equally: Tao Xue, Haifeng Jia, Meng Wang
* Corresponding authors, E-mail: xuejp@163.com; linjishan2006@163.com; yboduan@163.com

Abstract

The tuber of *Pinellia ternata* is broadly used in traditional herbal medicines in Asian countries. Here, we report a chromosome-level genome sequence of *P. ternata*. The genome of diploid *P. ternata* was 2.08 Gb and assembled into 13 pseudo-chromosomes containing 34,342 genes. *P. pedatisecta* is the closest-related species with a full genome sequence, and their divergence began approximately 11.55 million years ago. Comparative transcriptome analysis on various tissues suggested the enrichment of genes involved in phenylpropanoid biosynthesis and starch and sucrose metabolism in *P. ternata* tuber. The candidate genes of ephedrine biosynthesis in the phenylpropanoid pathway were identified. At least one gene for each synthase of ephedrine biosynthesis was predominantly expressed in the tuber. Notably, all four phenylalanine ammonia lyase genes were predominately expressed in tuber tissue. A series of genes involved in oxalate metabolism were found to be highly expressed in tubers, contributing to the high accumulation of oxalate in tubers as well as the formation of acrid raphide via reaction with calcium ions. There are 14 lectin genes in the *P. ternata* genome, which were all highly expressed in the tuber, explaining the acrid raphide formation. These findings provide new insight into ephedrine biosynthesis and acrid raphide formation.

Citation: Xue T, Jia H, Wang M, Zhang Y, Liu X, et al. 2024. A chromosome-level *Pinellia ternata* genome assembly provides insight into the evolutionary origin of ephedrine and acrid raphide formation. *Medicinal Plant Biology* 3: e013 https://doi.org/10.48130/mpb-0024-0012

Introduction

Pinelliae Rhizoma, the dried tuber of *Pinellia ternata* (Thunb.) Breit. (Araceae), is a famous bulk materia medica broadly used in traditional herbal medicines in Asian countries[1–3]. Particularly, Xiaoqinglong Tang (Shoseiryuto), Banxia Houpu Tang (Hangekobokuto), and Wenjing Tang (Unkeito) are the classical formulas widely used in Traditional Chinese Medicine (TCM) and Japanese Kampo medicine; the dried tuber of *P. ternata* is the key component of these medicines[4]. Moreover, other formulas, such as Banxia Shumi Tang, are also broadly applied in TCM. Medicines made using *P. ternata* primarily have antiemetic, expectorant, sedative, antipyretic, and styptic effects[5–8]. These medicinal functions are attributed to various bioactive compounds, including alkaloids, iridoid glycosides, and organic acids in the tuber, as reported in several pharmacological studies[9–12]. Of the compositions, alkaloids are deemed as the core component of the medicinal efficacy of *P. ternata* tubers[12–14].

The alkaloids present in *P. ternata* are guanosine, inosine, trigonelline, and ephedrine[15]. Among them, ephedrine is of particular interest due to its function as an antitussive[16,17]. Ephedrine, the key bioactive component of *P. ternata*, accumulates predominantly in tubers[18], and the ephedrine content in tubers typically ranges from 6 to 14 $\mu g \cdot g^{-1}$ dry weight[19]. In addition to expanding the cultivation acreage, the breeding or selection of *P. ternata* with high medicinal components is an important approach to meet the market demand. Understanding the biosynthesis pathway of ephedrine may allow for an increase in its accumulation and a decrease in the dose of *P. ternata* tuber. Previous studies have investigated the biosynthesis pathway of ephedrine in *P. ternata*. Specifically, Zhang et al.[20] and Duan et al.[14] conducted a transcriptome analysis and external elicitation to construct the biosynthetic pathway of ephedrine in *P. ternata*. However, our understanding of the pathway was based on transcriptome data, which is insufficient to provide complete candidate genes for each biosynthatase. In addition, the final steps from 1-phenylpropane-1,2-dione to ephedrine remain unclear.

In addition to its medically relevant alkaloids, *P. ternata* has an acrid taste and is defined as 'toxic' in The Chinese Pharmacopoeia[21]. The acridity of *P. ternata* tubers, as well as the acridity of several other Araceae species is mainly attributed to

insoluble needle-like crystals called acrid raphides[22,23]. Recent studies showed that the raphides are composed of calcium oxalate, proteins, and polysaccharides[24]. Acrid raphides are calcium oxalate by a conjugated reaction between calcium and oxalic acid, and exogenous fertigation of calcium could enhance calcium oxalate[25]. The acridity of the *P. ternata* tuber can be explained by the calcium oxalate crystals that mediate the introduction of inflammatory proteins (lectins) by penetrating the mucous membrane[26]. Because of the presence of acrid raphides, proper processing *via* decocting with licorice, alum, or ginger is required while making traditional medicines using *P. ternata* tubers[4,27]. Processing, however, would lower alkaloid concentrations significantly in *P. ternata* tubers[28]. Though the composition of raphides is understood, it is currently unclear how acrid raphides form in Araceae species.

Though there are approximately 110 genera and 3,500 species in Araceae, only *Spirodela polyrhiza*[29], *Colocasia esculenta*[30], *Pinellia pedatisecta*[31], *Pistia stratiotes*[32] and *Amorphophallus konjac* K. Koch[33] have fully sequenced genomes. Moreover, *S. polyrhiza* and *P. stratiotes* are an aquatic plant, and *C. esculenta* and *A. konjac* are vegetable crops; therefore, their genome sequences may not be sufficiently compatible to provide a strong reference for *P. ternata*. *P. pedatisecta* is a widely used as traditional medicinal herbs[9]. However, *P. ternata* and *P. pedatisecta* belong to different species, which will be different in many aspects. The lack of genome information largely restricts molecular studies on *P. ternata*. Typically, *P. ternata* is a hexaploid or octoploid in production[34], increasing the complexity of genome sequencing. However, using a wild diploid accession, we sequenced the *P. ternata* genome and provided a chromosome-scale assembly by combining Illumina, Pacbio HiFi, and Hi-C sequencing technology. Further, the biosynthesis pathways for amphetamine-type alkaloids and acrid raphides were constructed. The data presented herein will provide valuable information for illustrating the molecular mechanisms controlling the biosynthesis of alkaloids and acrid raphides.

Materials and methods

Plant materials, library construction, and sequencing

Individual *P. ternata* Breit. plants that were collected from their wild habitat in Baojia Town, Yuexi County, Anhui Province, PR China (30°83′ N, 116°08′ E) and identified as diploids via flow cytometry, were planted in the greenhouse at Huaibei Normal University. High-quality genomic DNA was extracted from young leaves of *P. ternata* for genome sequencing. A paired end library was constructed and sequenced using the BioMarker Technology Company (BioMarker, Qingdao, China) and Illumina NovaSeq X platforms to generate short reads. The PacBio sequel II platform was used for Pacbio HiFi library constructing and sequencing, thus obtaining long reads. The Hi-C library was constructed using the same leaf samples as described previously and sequenced using Illumina NovaSeq X platforms. Hi-C-Pro was applied for quality control and the identification of valid interaction pairs.

Genome size estimation

Two methods were used to estimate the genome size of *P. ternata*: flow cytometry and K-mer counting. GenomeScope2.0[35] was used to estimate the characteristics of the *P. ternata* genome, including genome size, heterozygosity, and repetitiveness based on the K-mer method.

Genome assembly and quality assessment

Pacbio HiFi long reads were assembled using HiFiasm[36]. Redundant sequences in the initial assembly of *P. ternata* were removed using Khaper (https://github.com/lardo/khaper). NextPolish[37] was used to polish the genome assembly, with Illumina reads employed to generate the final contig-level assembly. ALL Hi-C was used to anchor the contigs to chromosomes based on the HiC reads, and JUICEBOX was used to correct assembly errors[38], allowing for the attainment of a chromosome-level genome. BUSCO (v5.2.2)[39] with OrthoDB (embryophyta_odb10), Illumina short reads and transcriptome data were mapped to the *P. ternata* genome with BWA-mem to assess the quality of the assembled genome. The Long-Terminal Repeat (LTR) Assembly Index (LAI)[40] was also used to evaluate the quality of the genome. The evaluation criteria were as follows: the genome assembly will be considered a draft genome if the LAI value was between 0 and 10, the genome assembly will be considered a reference genome if the LAI value was between 10 and 20, and the genome assembly will be taken as gold if the LAI value was greater than 20.

Gene prediction and functional annotation

GETA, a homology-based tool integrating three methods, was used to predict the protein-coding genes in the *P. ternata* genome (https://github.com/chenlianfu/geta). Genes were functionally annotated by searching against the eggNOG (Evolutionary Genealogy of Genes: Non-supervised Orthologous Groups) 5.0 database. The protein-coding sequences were submitted to the EggNOGmapper software online with the default parameter settings.

Repetitive sequence annotation

RepeatMasker (v4.0.9) and RepeatModeler (v1.0.8)[41] were used to identify repetitive sequences in the *P. ternata* genome. RepeatModeler was used for *de novo* prediction and a library of consistent repetitive sequences was established. RepeatMasker was then used for comprehensive annotation by loading the repetitive sequence library. Unknown repeat sequences were further classified using TEclass[42]. Long-terminal repeat retrotransposons (LTR-RTs) were identified using LTR Finder (v1.0.2)[43] and LTRharvest.

Genome evolution and divergence time estimation

To conduct an evolutionary analysis, we collected the protein sequences of *P. ternata* and 13 other species: *Arabidopsis thaliana*, *Carica papaya*, *Oryza sativa*, *Brachypodium distachyon*, *Areca catechu*, *Wolffia australiana*, *Elaeis guineensis*, *Cocos nucifera*, *Colocasia esculenta*, *Amorphophallus konjac*, *Pinellia pedatisecta*, *Pistia stratiotes*, and *Zostera marina*. We also identified the gene clusters of these genomes using Orthofinder[44] with default settings. CAFE (v4.2.1)[45] was used to identify the gene families that underwent expansion or contraction in the 11 sequenced species.

Single-copy orthologous genes from the 14 species were extracted and the protein sequences were aligned using MAFFT (v7.307)[46]. RaxML (v 8.2.12)[47] was used to construct the maximum-likelihood phylogenetic tree. The MCMCTree program in the PAML package[48] was used to estimate the species divergence times using the divergence time between *A.*

thaliana and *C. papaya* (i.e., 68–72 million years ago) and the monocot and eudicot divergence time (i.e., 120–140 million years ago), as shown by *A. thaliana* and *O. sativa* as calibrators.

WGD event andsSynteny analysis

P. ternata, *A. konjac*, *P. stratiotes*, and *P. pedatisecta* belong to the subfamily of *Aroideae* of the Araceae family. To identify potential WGD events of the *P. ternata* genome, the WGD pipeline[49] was used to calculate the distribution of synonymous substitutions per synonymous site (Ks).

JCVI[50] with default parameters was used to identify collinear blocks between *P. ternata*, *A. konjac*, and *P. pedatisecta*, and a dot plot was drawn to confirm synteny between *P. ternata* and *P. pedatisecta* genomes.

Integrated genomic and transcriptomic analysis of ephedrine and oxalate biosynthesis-associated genes

Total RNA was extracted from tubers during different growth periods (young (Y), medium (M), and old (O)), as well as roots, stems, leaves, and flowers using the Plant RNA Kit following the manufacturer's instructions. Three biological replicates were conducted for each tuber sample. The isolated RNA was assessed using a NanoDrop Qubit 2.0 Fluorometer. High-quality RNA was sent to Berry Genomics (www.berrygenomics.com) for library construction and sequencing.

The sequencing reads were aligned to the assembled genome using HISAT2[51]. Stringtie2[52] was used to calculate the expression levels of all genes through fragments per kilobase of transcript per million mapped fragments (FPKM). DeSeq2[53] was used to analyze the significantly differentially expressed genes (DEGs) with a false discovery rate (FDR) of ≤ 0.05 and an absolute log2 (fold change) of ≥ 1 as the threshold. DEGs were clustered into different expression files using R package ClusterGVis[54]. iTAK (v 1.7)[55] was used to identify the transcription factors (TFs) in the genome and classify them based on the PlnTFDB and PlantTFDB databases.

After removing redundant sequences, we identified genes potentially involved in the biosynthesis pathway of ephedrine and oxalate using BLASTP (E-value =1e^{-5})[56].

Results

Genome sequencing, assembly, and annotation

According to the flow cytometry results, the estimated genome size of the *P. ternata* genome (2n = 2x = 26) was ~1.9 Gb (Supplemental Fig. S1, Supplemental Table S1). Based on K-mer counting, a similar genome size was obtained (Supplemental Fig. S2, Supplemental Table S2). The heterozygosity and repetitiveness of the *P. ternata* genome were 2.51% and 76.2% (Supplemental Table S2) respectively, indicating that the *P. ternata* genome was a complex diploid genome with high heterozygosity and high repetition. A total of 101.61 Gb Illumina short reads (~48.8×), 65.78 Gb Pacbio HiFi long reads (~31.6×), and 216.46 Gb Hi-C reads (~104.1×) were obtained for the assembly of the *P. ternata* genome (Supplemental Tables S3–S5). The final assembled sequence was 2.08 Gb with anchoring on 13 pseudo-chromosomes that ranged from 117.77 Mb to 189.55 Mb long (Fig. 1, Supplemental Table S6 & S7). The total length of the 13 pseudo-chromosomes accounted for 95.46% of the total *P. ternata* genome sequence (Supplemental Fig. S3, Supplemental Table S7). The BUSCO score was 92.60% (1,495), and 84.8% (1,369) of the genes were considered to be complete and single copy (Supplemental Table S8). The LAI value was 21.87 (LAI > 20) (Supplemental Table S9).

We predicted 34,342 protein-coding genes in the *P. ternata* genome by integrating homology-based, *de novo*, and transcriptome-based methods. Among the predicted genes, 32,140 (93.6%) genes were functionally annotated using EggNOGmapper. BUSCO (v5.2.2) analysis found that 1,332 (82.5%) were identified as complete (Supplemental Table S8).

Through a combination of *de novo* and homology-based approaches, 86.02% of the *P. ternata* genome assembly was annotated as repeat sequences, including 72.36% of retrotransposons and 10.9% transposons (Supplemental Table S10). The LTR retrotransposons accounted for the largest proportion (62.59%) of retrotransposons and their length was approximately 1,243.8 Mb, including 342.08 Mb of Ty1/Copia, 722.48 Mb of Ty3/Gyspy, and 179.24 Mb of other retrotransposons, which accounted for 17.21%, 36.36%, and 9.02% of genome assembly (Supplemental Table S10), respectively.

Genome evolution and expansion/contraction of gene families

Phylogenetic trees were constructed with 22 single-copy gene families (Supplemental Table S11), which showed that *P. ternata* and *P. pedatisecta* were sisters to each other and diverged approximately 11.55 mya. *P. pedatisecta* was the closest genome to *P. ternata* (Fig. 2a). Compared with other 13 species indicated that 1,633 gene families were expanded and 4,096 gene families were contracted in *P. ternata* (Supplemental Table S12). Additionally, 1,112 gene families containing 6,654 genes underwent rapid evolution (Supplemental Table S11). GO and KEGG enrichment analyses were performed and the results revealed that GO analysis enriched in the processes of DNA biosynthesis, RNA-dependent DNA replication and sphingoid biosynthetic and metabolic process (Supplemental Fig. S4a), while KEGG analysis enriched in pathways of Sphingolipid metabolism, Photosynthesis, and Isoquinoline alkaloid biosynthesis (Supplemental Fig. S4b).

The gene family clustering analysis of 14 plant species showed a total of 482,019 genes were clustered into 20,472 gene families (Supplemental Table S11). Among them, a total of 13,200 gene families comprising 28,094 genes were identified in *P. ternata* genome (Supplemental Table S13). The *P. ternata* genome assembly was also compared with *A. konjac*, *P. pedatisecta*, *P. stratiotes, C. esculenta*,, and *W. australiana*. In total, 8,051 gene families common to these six species were identified, and 373 gene families containing 818 genes were specific to *P. ternata* (Fig. 2b). GO enrichment analysis indicated that the *P. ternata*-specific genes were mainly enriched in the organic substance biosynthetic process, cellular nitrogen compound biosynthetic process, and cellular biosynthetic process. These enriched genes might be associated with the biosynthetic process of *P. ternata* (Supplemental Fig. S5a). KEGG enrichment analysis showed that *P. ternata*-specific genes were mainly enriched for biosynthesis of secondary metabolites, Riboflavin metabolism, Carotenoid biosynthesis, and valine, leucine and isoleucine degradation pathways (Supplemental Fig. S5b).

WGD events and synteny analysis

Whole-genome duplication (WGD) events were investigated in *P. ternata*, *A. konjac*, *C. esculenta*, and *P. pedatisecta* genomes. *P. ternata* with the other three species belonging to *Aroideae*

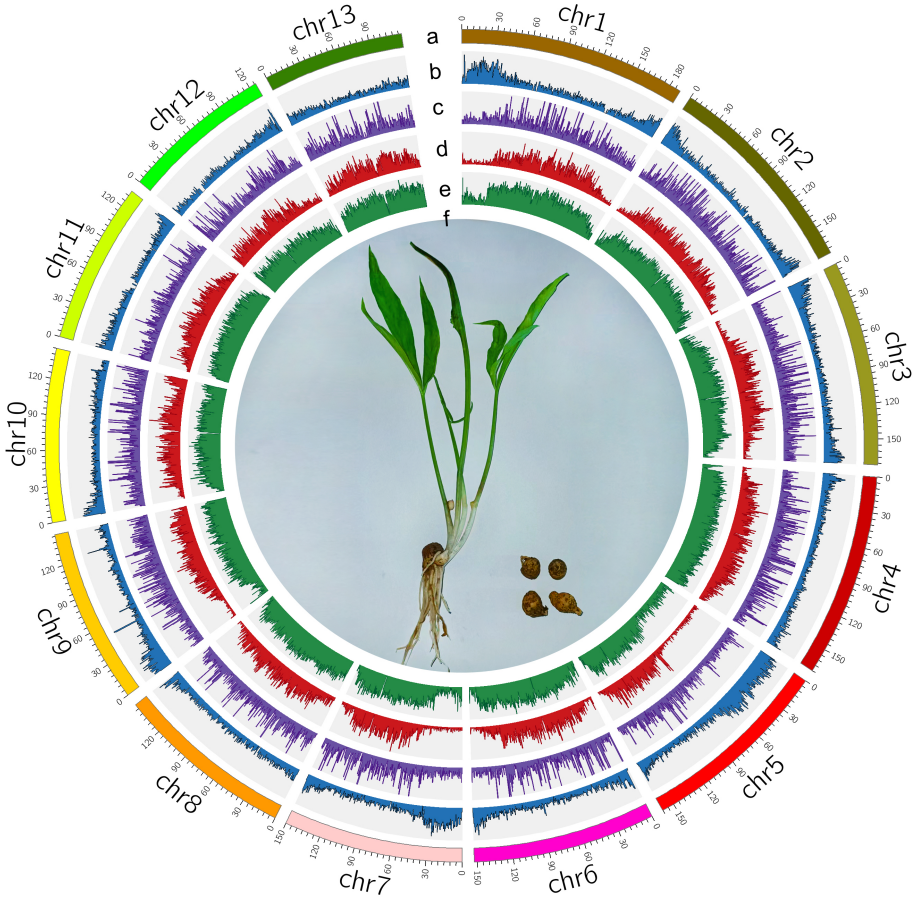

Fig. 1 Genome features of *Pinellia ternata*. (a) Chromosomes (chr01-Mbchr13); (b) Gene density; (c) DNA transposable elements; (d) LTR/Copia; (e) LTR/Gypsy; (f) *P. ternata*.

shared one ancient WGD event. The distribution of Ks values in *P. ternata* showed two peaks at about 0.05 and 0.25 (Fig. 2c), indicating that *P. ternata* experienced a WGD event twice and that the most recent WGD event occurred independently after *P. ternata* and *P. pedatisecta* diverged.

A synteny analysis was conducted with the *P. ternata, A. konjac* and *P. pedatisecta* genomes. A total of 13,182 syntenic gene pairs across 379 syntenic blocks between the *P. ternata* and *P. pedatisecta* genomes were identified. A total of 11,549 gene pairs across 551 syntenic blocks between the *P. ternata* and *A. konjac* genomes were identified. Fragmented conserved synteny was identified for the comparative genome structure between *P. ternata* and *P. pedatisecta* (Fig. 2d, Supplemental Fig. S6). In addition, a 1:1 syntenic depth ratio was observed in the *P. ternata* – *P. pedatisecta* and *P. pedatisecta* – *P. ternata* genomes. For most collinear regions, one chromosome of *P. ternata* corresponded to one chromosome of *P. pedatisecta*. For example, Ptchr01, Ptchr02, Ptchr03, Ptchr05, Ptchr06, and Ptchr08 of *P. ternata* corresponded to Ppchr06, Ppchr03, Ppchr01, Ppchr10, Ppchr04, and Ppchr08 of *P. pedatisecta,* respectively. The collinearity of *P. ternata* and *P. pedatisecta* is obviously better than that of *P. ternata* and *A. konjac* (Fig. 2d).

Transcriptome analysis of *P. ternata*

To gain a better understanding of the medicinal components of *P. ternata*, a detailed transcriptome analysis was performed on *P. ternata* tubers of different developmental stages. We compared the tuber of *P. ternata* in three different developmental stages and that DEGs were identified from O vs Y, and M vs Y comparisons, with 1,394 and 866 DEGs, respectively (Supplemental Table S14). Among the DEGs, 1,039 were specific to O vs Y, and 511 were specific to M vs Y, while 355 DEGs were common in M vs Y and O vs Y (Supplemental Fig. S7).

Next, ClusterGVis was employed to analyze the expression heatmap of the DEGs, and these DEGs among the different developmental stages were clustered into eight expression profiles. A large set of genes (n = 874, clusters 1, 2, 3) expressed at high levels in the Y stage and 386 genes (clusters 1, 2, 3) expressed highly in the M stage, while genes in clusters 6, 7 and 8 (n = 645) expressed at high levels in the O stage (Fig. 3). GO and KEGG analysis were employed to further annotate these genes with high expression levels in different stages. For GO analysis, genes in clusters 1, 2, and 3 were mainly enriched in biological process, including cell cycle process, cell differentiation, single-organism developmental process (Supplemental Fig. S8a). Genes in clusters 4, and 5 had enrichment in biological processes that respond to gibberellin and abiotic stimuli (Supplemental Fig. S8b), while genes in clusters 6, 7, and 8 were mainly enriched in lignin biosynthetic and metabolic processes, phenylpropanoid biosynthetic and metabolic process (Supplemental Fig. S8c). For KEGG analysis, genes with high expression levels in different stages were all enriched in the phenylpropanoid biosynthesis pathways (Fig. 3). Except for enrichment in phenylpropanoid biosynthesis pathways, genes with

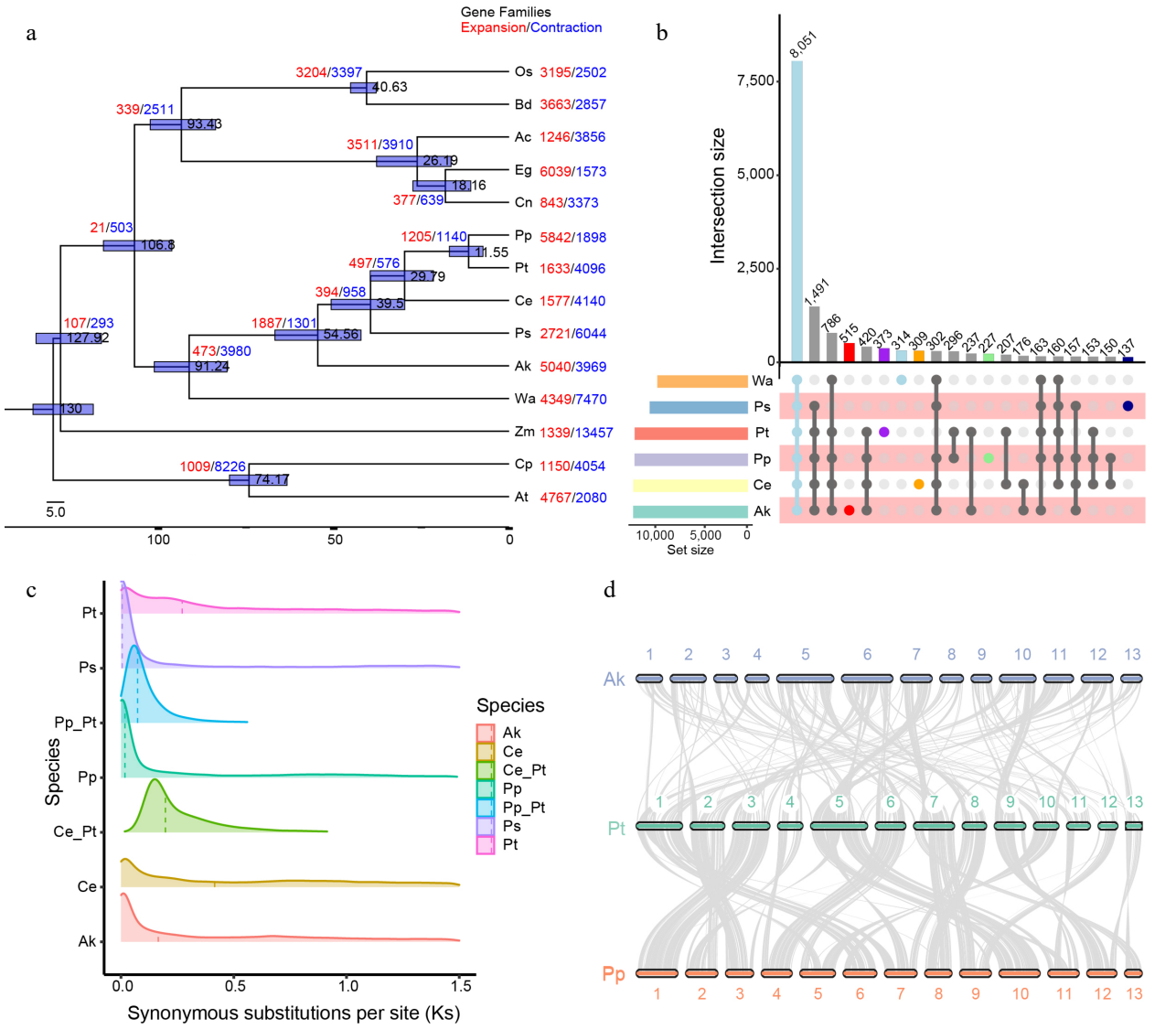

Fig. 2 Evolution and gene family analysis of *Pinellia ternata* and other representative plant genomes. (a) Phylogenetic analysis of the *P. ternata* genome and expansions and contractions of gene families. Gene family expansions and contractions are indicated by the numbers in red and blue, respectively. (b) The distribution of shared and specific gene families among six species. Upper diagram showing the distribution of shared gene families among six species. Lower diagram showing gene family complement comparisons among *Wolffia australiana*, *Amorphophallus konjac*, *Pinellia pedatisecta* and *Pistia stratiotes*, *Colocasia esculenta*, and *P. ternata*. There were 8,051 common gene families and 373 *P. ternata*-specific gene families. (c) Ks distribution among *P. ternata* and three other species. Different color indicates Ks distribution within and between genomes. (d) Synteny between *P. ternata*, *Amorphophallus konjac* and *Pinellia pedatisecta*. The colored line connects matched gene pairs. The numbers indicate the corresponding chromosomes in each species, Pt, Pp and Ak represents *P. ternata*, *P. pedatisecta* and *A. konjac*, respectively.

high expression levels in the Y stage were also enriched in the pathways of pentose and glucuronate interconversions, amino sugar and nucleotide sugar metabolism and starch and sucrose metabolism. Those with high expression levels in the M stage were mainly enriched in tropane, piperidine, and pyridine alkaloid biosynthesis, ascorbate and aldarate metabolism, ubiquinone, and other terpenoid-quinone biosynthesis pathways. The high expression genes in the O stage were mainly enriched in stilbenoid, diarylheptanoid, and gingerol biosynthesis, flavonoid biosynthesis, and plant hormone signal transduction pathways. These combined results revealed that many secondary metabolites are produced during the different growth stages of tubers. Particularly, starch and polysaccharides may be synthesized in the Y stage, and alkaloids are mainly accumulated in the M stage, while stilbenoid, diarylheptanoid and gingerol may be generated with high production in the O stage of tuber.

The KEGG metabolic pathways of DEGs of different clusters were further investigated. Interestingly, genes in the cluster 6 mostly did not express in the Y and M stages but expressed highly in the O stage, and they were enriched in the pathways, including pentose and glucuronate interconversions, plant hormone signal transduction and protein processing in endoplasmic reticulum, suggesting that plant hormones may play a very important role in the tuber growth from the M to O stage.

The TFs in *P. ternata* were further annotated. A total of 1,564 TFs classified into 68 families were identified, which is approximately 3.5% of the total number of coding genes

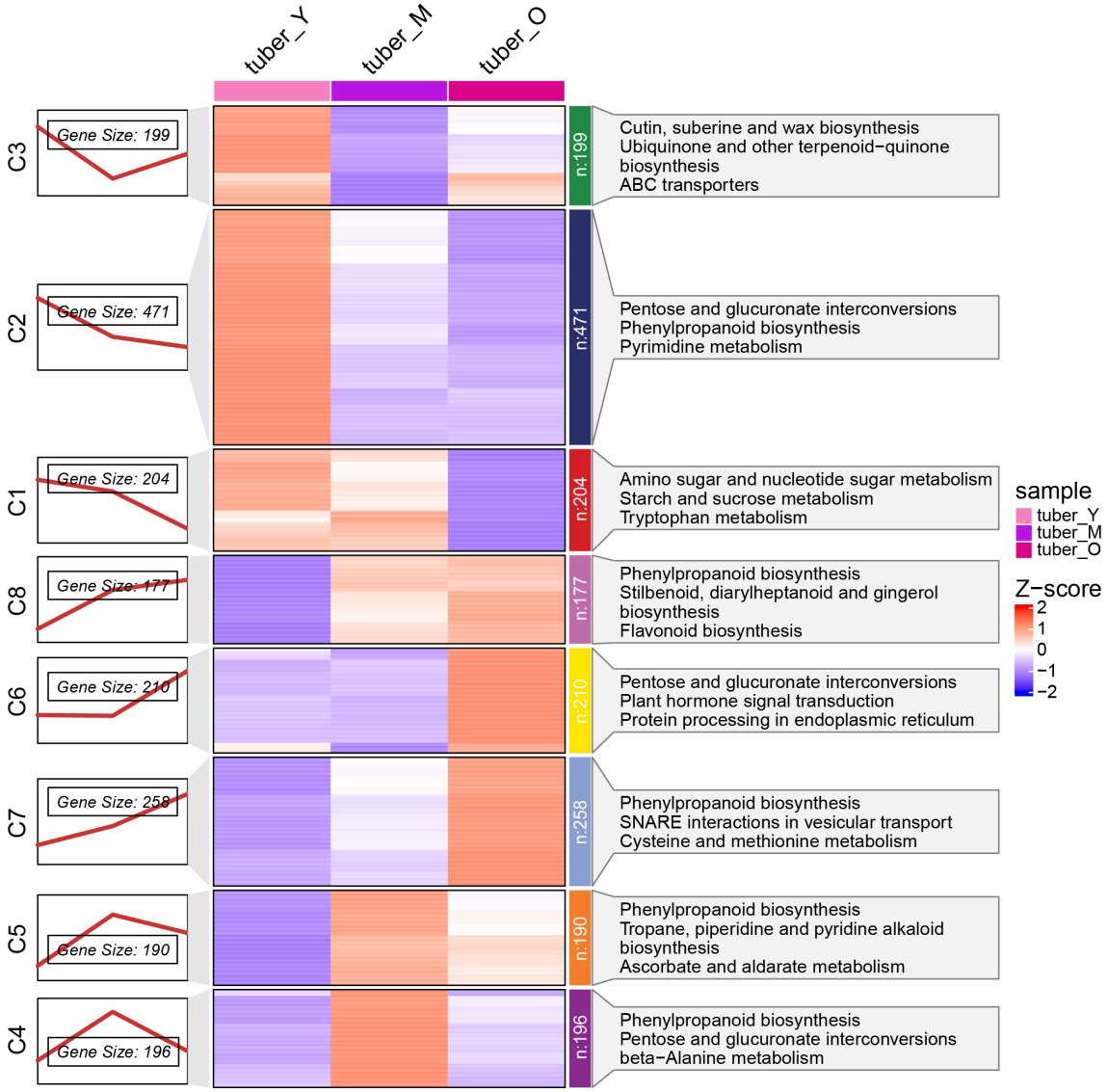

Fig. 3 Different gene regulatory landscape in different development stages of *Pinellia ternata* tuber. (a) Eight expression profile types of DEGs by K-Means clustering, (b) heatmap analysis of all DEGs in eight clusters, (c) pathway enrichment of DEGs in eight clusters.

(Supplemental Table S15). The bHLH family accounted for the largest proportion of coding genes (7.29%), followed by MYB (6.78%), NAC (6.52%), and C2H2 (6.07%). Of all the TFs, 128 were identified as DEGs, which accounts for 8.2% of all TFs and 6.7% of all DEGs. Comparing the three growth stages of tubers, the number of highly expressed TFs were most abundant in the Y stage, followed by O and M stages (Supplemental Fig. S9). The results indicate that these TFs play important roles in regulating tuber growth and development, as well as the bioactive constituent biosysnthesis during this period.

Analysis of key genes involved in ephedrine biosynthesis

The following candidate genes involved in the ephedrine biosynthesis pathway were identified based on the functional annotation and homology search: four phenylalanine ammonia lyase (PAL) genes, two cinnamate: CoA ligase (*CNL*) genes, two cinnamoyl-CoA hydratase-dehydrogenase (*CHD*) genes, six 3-hydroxyisobutyryl-CoA hydrolase (*CHY*) genes, three aldehyde oxidases 4 (*AO4*) genes, two 3-ketoacyl-CoA thiolase (*KAT*) genes, three benzaldehyde dehydrogenase (*BALDH*) genes, five benzoate-CoA ligase (*BL*) genes, three ThDP-dependent pyruvate decarboxylase (*ThPDC*) genes, five acetolactate synthase (*AHAS*) genes, and 13 N-methyltransferase genes (Fig. 4). The candidate genes that were predominantly expressed in tuber compared with the other tissues were as follows: *PtPAL PtGene58461*, *PtCNL PtGene57555*, *PtCHD PtGene34219*, *PtKAT PtGene49630*, *PtBL PtGene38901*, *PtCHY PtGene12715*, *PtBALDH PtGene47180*, *PtAAO4 PtGene00280*, and *PtAHAS PtGene07361*. It is notable that all four *PtPAL* genes were highly expressed in the root and tuber tissues at different developmental stages. It is possible that the *PtPAL* genes may be important for allocating the metabolic flux in the underground organs, further contributing to the accumulation of ephedrine in the tuber. For the *PtNMTs*, phylogenetic analysis, and expression profiles suggested both *PtGene15192* and *PtGene36030* expressed highly in tuber and had a close genetic relationship with known NMTs from other species (Supplemental Figs S10 & S11). Since NMT is the final key enzyme of ephedrine synthesis,

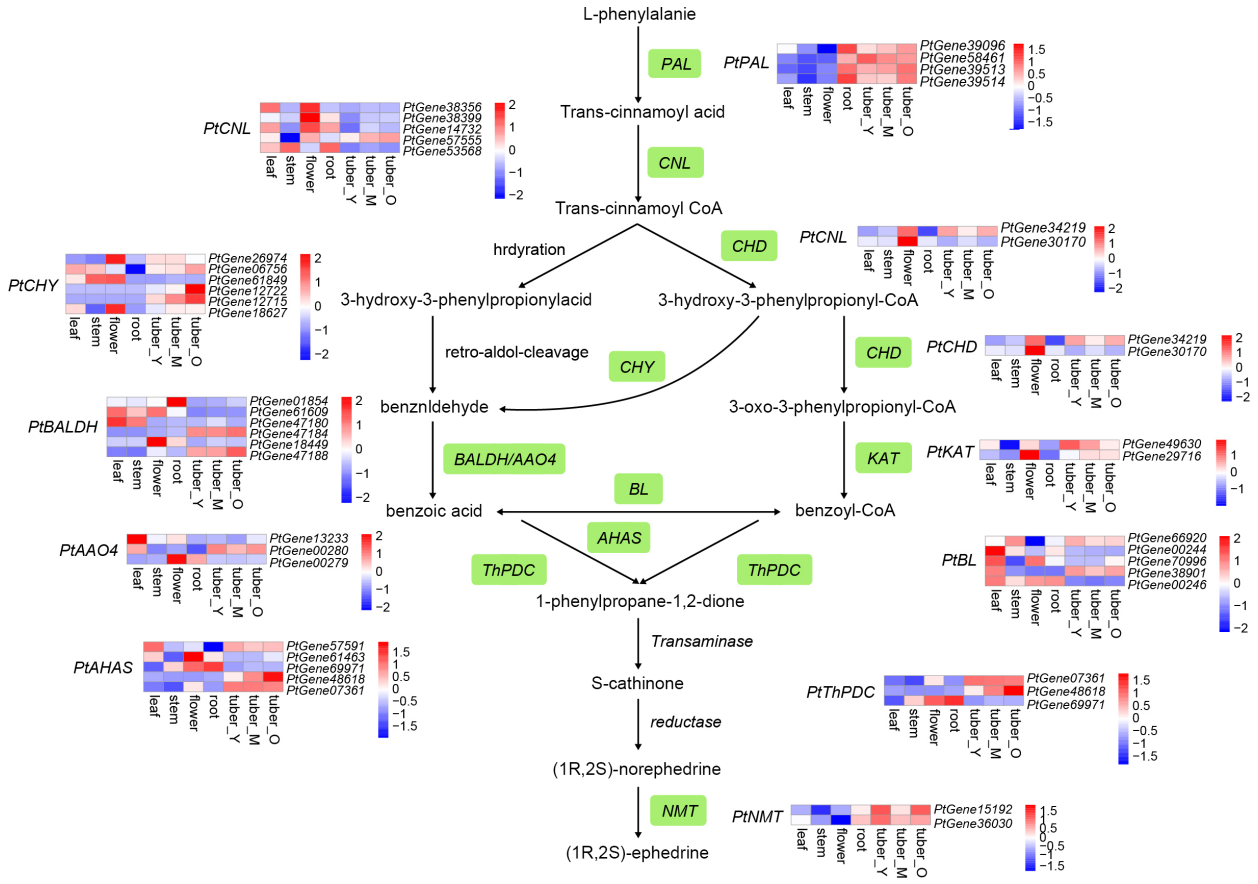

Fig. 4 Simplified representation of the ephedrine biosynthetic pathway. Top hits for pathway genes identified by blast are highlighted in green. The expression value for each gene is indicated in color on a log2 (FPKM + 1) scale for seven tissue types: leaf, stem, flower, root, young tuber (tuber_Y), medium tuber (tuber_M), and old tuber (tuber_O).

modulating its expression may increase the ephedrine content in *P. ternata* tubers.

Analysis of candidate genes involved in oxalate biosynthesis and the formation of acrid raphides

The candidate genes involved in oxalate biosynthesis were determined, and the putative oxalate biosynthesis pathways were outlined. Oxalate could be generated either from glyoxylate and oxaloacetate in the tricarboxylic acid (TCA) cycle or from the degradation of ascorbate[57]. The expression profile revealed the most probable candidate genes: *PtACO* PtGene58911, *PtGLO* PtGene39926, and *PtMDH PtGene24599* in the TCA cycle and *PtAPX PtGene48581* in the ascorbate degradation pathway (Fig. 5).

Moreover, 14 candidate genes responsible for lectin production were identified, and all were primarily expressed in the tuber (Fig. 5). The expression pattern of lectins may contribute to the formation of acrid raphides in *P. ternata*. The expression of the candidate genes related to oxalate biosynthesis, together with the lectin genes, may explain the formation of acrid raphides in *P. ternata* tubers.

Discussion

Genomically-relevant data is especially limited for studies on the bioactive compounds' biosynthesis mechanism in *P. ternata*. In the present study, a diploid accession of *P. ternata* was used for genome sequencing and transcriptome analysis, allowing us to produce a high-quality (LAI > 20) chromosome-level genome assembly of this plant species. Further, the genes involved in ephedrine biosynthesis and acrid raphide formation were identified and found to be highly expressed in the tuber.

As a phenylpropylamino alkaloid, ephedrine is generated in the L-phenylalanine (Phe) pathway, with Phe-derived benzoic acid as an intermediate. The reaction is initiated with a phenylalanine ammonia lyase (PAL) for deamination of phenylalanine to cinnamic acid[58]. Two possible pathways, i.e., β-oxidative (peroxisome) and non-β-oxidative (cytosol), have been proposed for benzoic acid biosynthesis[59]. In the β-oxidative routes, four synthases are involved, including the cinnamate: CoA ligase[60–62], cinnamoyl-CoA hydratase - dehydrogenase[59,63], 3-ketoacyl-CoA thiolase[64], and ThDP-dependent pyruvate decarboxylase[65]. In the non-β-oxidative pathway, two synthases, benzaldehyde dehydrogenase[66] and acetolactate synthase may be responsible for Phe side-chain shortening for benzoic acid biosynthesis[67,68]. Moreover, 3-hydroxyisobutyryl-CoA hydrolase is an enzyme that may connect the two pathways by converting trans-cinnamoyl-CoA to benzaldehyde[69]. Through transcriptome analysis, the genes encoding the above-mentioned biosynthases were obtained in *P. ternata*[20]. Nevertheless, none of them were primarily expressed in tubers, failing to explain the accumulation of ephedrine in tubers relative to other organs. In this study, by combining the tissue-specific transcriptome data with genome

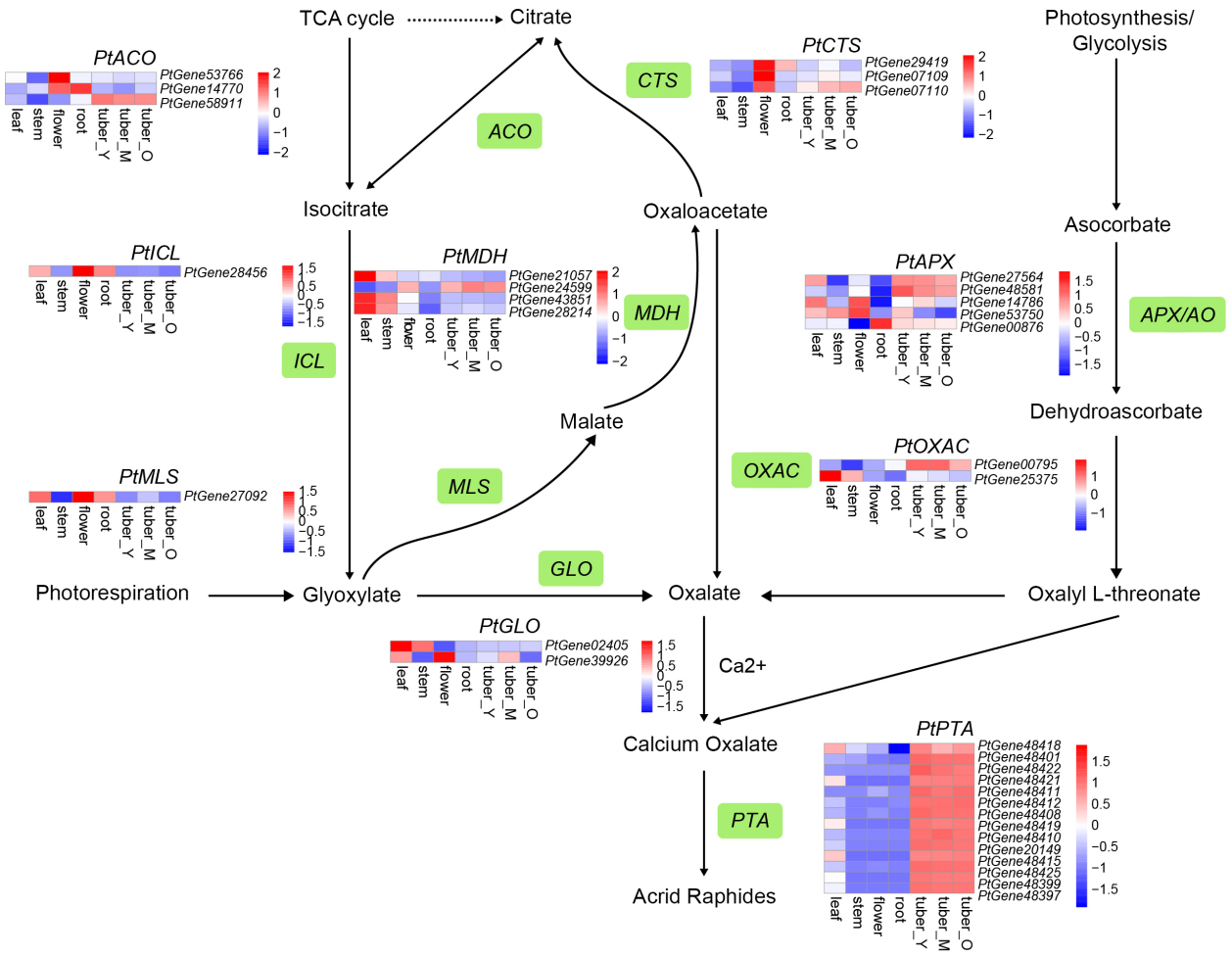

Fig. 5 Simplified representation of oxalate biosynthesis and acrid raphides formation. The enzyme names are as follows: glycolate oxidase (GLO); oxaloacetate acetylhydrolase (OXAC); malate synthase (MLS); malate dehydrogenase (MDH); citrate synthase (CTS); aconitase (ACO); isocitrate lyase (ICL); ascorbate peroxidase/ascorbate oxidase (APX/AO); lectin. Top hits for pathway genes identified by blast are highlighted in green. The expression value for each gene is indicated in color on a log2 (FPKM + 1) scale for seven tissue types: leaf, stem, flower, root, tuber_Y, tuber_M, and tuber_O.

data, at least one gene for each synthase was found predominantly in the tuber (Fig. 4). Secondary metabolites are usually accumulated and stored in the place of their synthesis[70]. Thus, the specific expression of the biosynthases is commonly correlated with their accumulation pattern. Unlike *Ephedra sinica*, which accumulates ephedrine in aboveground parts[71], *P. ternata* stores ephedrine in tubers. This indicates that the genes screened in this study are probable as the candidate genes for ephedrine biosynthesis in *P. ternata*. Moreover, the biosynthesis genes for the conversion of benzoic acid to ephedrine have not been completely discovered in *planta*, though a transaminase, a reductase, and an N-methyltransferase (NMT) may be involved based on their chemical structure[17,20] . Through phylogenetic and expression profiles, PtGene15192 and PtGene36030 are considered probable NMTs in *P. ternata*. Genetically manipulating their expressions may provide insight into breeding *P. ternata* germplasm with an increased ephedrine content.

In this study, we identified the metabolic pathway of oxalate and further outlined the formation mechanism of acrid raphide in *P. ternata*. The candidate gene *PtGene00795*, which was highly expressed in *P. ternata* tubers, may be important for oxalate accumulation *via* OXAC (catalyzing oxaloacetate to oxalate)[72]. There is just one candidate gene each for ICL (*PtGene28456*) and MLS (*PtGene27092*), and both have low expression levels in the tuber. This suggests that oxaloacetate rather than glyoxylate is the main precursor of oxalate via the TCA cycle, since oxaloacetate has another source from photosynthesis/glycolysis. In rice, however, glyoxylate is thought to be an efficient precursor for oxalate biosynthesis[73]. In other plants, both glyoxylate, and oxaloacetate were identified as the precursor[74]. Ascorbate degradation is another approach for oxalate formation[75]. Though some catalysis reactions remain clear, it is understood that ascorbate peroxidase/ascorbate oxidase (APX/AO) catalyzes the oxidation of ascorbate[76]. Among five candidate *APX/AO* genes, four were highly expressed in tubers (Fig. 5), suggesting that this enzyme is important for the specific biosynthesis of oxalate in tubers. Insecticidal, germicidal, and phlogogenic activities of *P. ternata* agglutinin (PTA) have been previously reported[77–79]. As shown in the *P. ternata* genome data, 14 *PTAs* were identified and all were highly expressed in tubers, which would facilitate the formation of acrid raphides, as well as the characterization of PTAs.

In conclusion, the genome of *P. ternata* was sequenced and assembled into 2.08 Gb in 13 pseudo-chromosomes. The expression profile analysis revealed the candidate genes for ephedrine biosynthesis and acrid raphide formation.

Author contributions

Project conception and coordination: Duan Y; experiments design: Xue T; supervision of genome sequencing and assembly: Lin J; genome annotation, synteny and gene family expansion/contraction analyses: Jia H, Wang M; transcriptional data analyses: Zhang Y, Liu X; karyotype analysis: Meng Z; plant materials and DNA extracts maintaining: Chao Q, Zhao F; draft manuscript preparation: Duan Y, Jia H; manuscript revise: Lin J, Xue T, Xue J. All authors reviewed the results and approved the final version of the manuscript

Data availability

All of the source data, statements of data availability and associated accession codes are available at: https://ngdc.cncb.ac.cn/search/?dbId=&q=PRJCA016404 with Bioproject ID: PRJCA016404. The *Pinellia ternata* genome assembly and annotation data are available at China National Center for Bioinformation (https://ngdc.cncb.ac.cn/gwh/Assembly/37791/show).

Acknowledgments

This work was supported by National Natural Science Foundation of China (82373993; 82274048), Excellent Scientific Research and Innovation Team of University in Anhui Province (2022AH010029), Natural Science Foundation Project of Fujian Province (2020J01593) and Hainan Provincial Natural Science Foundation of China (320RC711).

Conflict of interest

The authors declare that they have no conflict of interest.

Supplementary information accompanies this paper at (https://www.maxapress.com/article/doi/10.48130/mpb-0024-0012)

Dates

Received 8 March 2024; Revised 23 April 2024; Accepted 30 April 2024; Published online 18 July 2024

References

1. Zhang H, Zhang Z, Xiong Y, Shi J, Chen C, et al. 2021. Stearic acid desaturase gene negatively regulates the thermotolerance of *Pinellia ternata* by modifying the saturated levels of fatty acids. *Industrial Crops and Products* 166:113490
2. Bo C, Liu D, Yang J, Ji M, Li Z, et al. 2024. Comprehensive in silico characterization of NAC transcription factor family of *Pinellia ternata* and functional analysis of PtNAC66 under high-temperature tolerance in transgenic *Arabidopsis thaliana*. *Plant Physiology and Biochemistry* 208:108539
3. Yang JR, Cui WN, You Q, Liu MM, Liu X, et al. Transcriptome analysis reveals Long Non-Coding RNAs involved in shade-induced growth promotion in *Pinellia ternata*. *Frontiers in Bioscience-Landmark* 28(9): 202.
4. Fueki T, Tanaka K, Obara K, Kawahara R, Makino T, et al. 2020. The acrid raphides in tuberous root of *Pinellia ternata* have lipophilic character and are specifically denatured by ginger extract. *Journal of Natural Medicine* 74:722–31
5. Li Y, Li D, Chen J, Wang S. 2016. A polysaccharide from *Pinellia ternata* inhibits cell proliferation and metastasis in human cholangiocarcinoma cells by targeting of Cdc42 and 67 kDa Laminin Receptor (LR). *International Journal of Biological Macromolecules* 93:520–25
6. Xu JY, Dai C, Shan JJ, Xie T, Xie HH, et al. 2018. Determination of the effect of *Pinellia ternata* (Thunb.) Breit. on nervous system development by proteomics. *Journal of Ethnopharmacology* 213:221–29
7. Zhao JL, Li Q, Ding YY, Gu XY, Feng WW, et al. 2018. Sedative activity of the chemical constituents of *Rhizoma pinelliae* praeparatum. *Chemistry of Natural Compounds* 54:215–17
8. Mao R, He Z. 2020. *Pinellia ternata* (Thunb.) Breit: a review of its germplasm resources, genetic diversity and active components. *Journal of Ethnopharmacology* 263:113252
9. Ji X, Huang B, Wang G, Zhang C. 2014. The ethnobotanical, phytochemical and pharmacological profile of the genus *Pinellia*. *Fitoterapia* 93:1–17
10. Lee JY, Park NH, Lee W, Kim EH, Jin YH, et al. 2016. Comprehensive chemical profiling of *Pinellia* species tuber and processed *Pinellia* tuber by gas chromatography–mass spectrometry and liquid chromatography– atmospheric pressure chemical ionization–tandem mass spectrometry. *Journal of Chromatography A* 1471:164–77
11. Tang D, Yan R, Sun Y, Kai G, Chen K, et al. 2020. Material basis, effect, and mechanism of ethanol extract of *Pinellia ternata* tubers on oxidative stress-induced cell senescence. *Phytomedicine* 77:153275
12. Xue T, Xiong Y, Shi J, Chao Q, Zhu Y, et al. 2021. UHPLC-MS-based metabolomic approach for the quality evaluation of *Pinellia ternata* tubers grown in shaded environments. *Journal of Natural Medicines* 75:1050–57
13. Wu YY, Huang XX, Zhang MY, Zhou L, Li DQ, et al. 2015. Chemical constituents from the tubers of *Pinellia ternata* (Araceae) and their chemotaxonomic interest. *Biochemical Systematics and Ecology* 62:236–40
14. Duan Y, Zhang H, Meng X, Huang M, Zhang Z, et al. 2019. Accumulation of salicylic acid-elicited alkaloid compounds in *in vitro* cultured *Pinellia ternata* microtubers and expression profiling of genes associated with benzoic acid-derived alkaloid biosynthesis. *Plant Cell, Tissure and Organ Culture* 139:317–25
15. Liu YH, Liang ZS, Chen B, Yang DF, Liu JL. 2010. Elicitation of alkaloids in *in vitro* PLB (protocorm-like body) cultures of *Pinellia ternata*. *Enzyme and Microbial Technology* 46:28–31
16. Moustafa AA, Hegazy MA, Mohamed D, Ali O. 2018. Novel approach for the simultaneous determination of carbinoxamine maleate, pholcodine, and ephedrine hydrochloride without interference from coloring matter in an antitussive preparation using smart spectrophotometric methods. *Journal of Aoac International* 101(2):414–26
17. Morris JS, Groves RA, Hagel JM, Facchini PJ. 2018. An *N*-methyltransferase from *Ephedra sinica* catalyzing the formation of ephedrine and pseudoephedrine enables microbial phenylalkylamine production. *Journal of Biological Chemistry* 293:13364–76
18. Xu WF, Zhang BG, Li M, Liu GH. 2007. Determination of ephedrine in rhizoma Pinelliae, rhizoma *Typhonii flagelliformis* and their processed products by HPLC. *Lishizhen Medicine and Materia Medica Research* 18(4):884–85
19. Wen Q, Zhang Y, Zhang J, Zhao Z. 2016. Simultaneous determination of 6 organic acids, 3 nucleosides, and ephedrine in *Pinellia ternata* by HPLC. *Journal of Chinese Pharmaceutical Sciences* 25(12):906–13
20. Zhang GH, Jiang NH, Song WL, Ma CH, Yang SC, et al. 2016. *De novo* sequencing and transcriptome analysis of *Pinellia ternata*

identify the candidate genes involved in the biosynthesis of benzoic acid and ephedrine. *Frontiers in Plant Science* 7:1209
21. National Pharmacopoeia Committee. 2020. *Pharmacopoeia of the People's Republic of China*. Beijing: China Medical Science and Technology Press.
22. Wu H, Li W, Han H, Ji R, Ye D. 1999. Studies on stimulating components of raw *Pinellia ternata* (Thunb.) (Banxia). *China Journal of Chinese Materia Medica* 24:725−30
23. Zhong LY, Wu H, Zhang KW, Wang QR. 2006. Study on irritation of calcium oxalate crystal in raw *Pinellia ternata*. *China Journal of Chinese Materia Medica* 31:1706−10
24. Ge XY, Wu H. 2010. Analysis of the composition of poisonous raphides in Araceae plant. *Chinese Journal of Pharmaceutical Analysis* 30(2):190−93
25. Smith KT, Shortle WC, Connolly JH, Minocha R, Jellison J. 2009. Calcium fertilization increases the concentration of calcium in sapwood and calcium oxalate in foliage of red spruce. *Environmental and Experimental Botany* 67:277−83
26. Paull ER, Tang CS, Gross K, Uruu G. 1999. The nature of the taro acridity factor. *Postharvest Biology and Technology* 16(1):71−78
27. Su T, Zhang WW, Zhang YM, Cheng BCY, Fu XQ, et al. 2016. Standardization of the manufacturing procedure for Pinelliae Rhizoma praeparatum cum zingibere et alumine. *Journal of Ethnopharmacology* 193:663−69
28. Zhai XY, Zhang L, Li BT, Feng YL, Xu GL, et al. 2019. Discrimination of toxic ingredient between raw and processed *Pinellia ternata* by UPLC/Q-TOF-MS/MS with principal component analysis and t-test. *Chinese Herbal Medicines* 11:200−8
29. An D, Zhou Y, Li C, Xiao Q, Wang T, et al. 2019. Plant evolution and environmental adaptation unveiled by long-read whole-genome sequencing of *Spirodela*. *Proceedings of the National Academy of Sciences of the United States of America* 116:18893−99
30. Yin J, Jiang L, Wang L, Han X, Guo W, et al. 2021. A high-quality genome of taro (*Colocasia esculenta* (L.) Schott), one of the world's oldest crops. *Molecular Ecology Resources* 21:68−77
31. Qian Z, Ding J, Li Z, Chen J. 2022. The high-quality *Pinellia pedatisecta* genome reveals a key role of tandem duplication in the expansion of its agglutinin genes. *Horticulture Research* 10(3):uhac289
32. Qian Z, Li Y, Yang J, Shi T, Li Z, et al. 2022. The chromosome-level genome of a free-floating aquatic weed *Pistia stratiotes* provides insights into its rapid invasion. *Molecular Ecology Resources* 22(7):2732−43
33. Gao Y, Zhang Y, Feng C, Chu H, Feng C, et al. 2022. A chromosome-level genome assembly of *Amorphophallus konjac* provides insights into konjac glucomannan biosynthesis. *Computational and Structural Biotechnology Journal* 20:1002−11
34. Lu J, Liu JN, Sarsaiya S, Duns GJ, Han J, et al. 2020. Phenotypic and transcriptomic analysis of two *Pinellia ternata* varieties T2 line and T2Plus line. *Scientific Reports* 10:4614
35. Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and smudgeplot for reference-free profiling of polyploid genomes. *Nature Communications* 11:1432
36. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nature Methods* 18:170−75
37. Hu J, Fan J, Sun Z, Liu S. 2020. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* 36:2253−55
38. Zhang X, Zhang S, Zhao Q, Ming R, Tang H. 2019. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nature Plants* 5:833−45
39. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO Update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution* 38:4647−54
40. Ou S, Chen J, Jiang N. 2018. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Research* 46(21):e126
41. Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols Bioinformatics* 25:4.10.1−4.10.14
42. Hoede C, Arnoux S, Moisset M, Chaumier T, Inizan O, et al. 2014. PASTEC: An automatic transposable element classification tool. *Plos ONE* 9:e91929
43. Xu Z, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Research* 35:W265−W268
44. Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* 20:238
45. Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW. 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Molecular Biology and Evolution* 30:1987−97
46. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30:772−80
47. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312−13
48. Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24(8):1586−91
49. Zwaenepoel A, Van de Peer Y. 2019. wgd—simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics* 35:2153−55
50. Tang H, Krishnakumar V, Li J. 2015. JCVI: JCVI utility libraries (v0.5.7). Zenodo. https://doi.org/10.5281/zenodo.31631
51. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology* 37:907−15
52. Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, et al. 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biology* 20:278
53. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15(12):550
54. Zhang J. 2022. ClusterGVis: One-step to cluster and visualize gene expression Matrix. https://github.com/junjunlab/ClusterGVis.
55. Zheng Y, Jiao C, Sun H, Rosli HG, Pombo MA, et al. 2016. iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Molecular Plant* 9:1667−70
56. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, et al. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421
57. Smirnoff N. 2018. Ascorbic acid metabolism and functions: a comparison of plants and mammals. *Free Radical Biology and Medicine* 122:116−29
58. Howles PA, Sewalt VJH, Paiva NL, Elkind Y, Bate NJ, et al. 1996. Overexpression of L-phenylalanine ammonia-lyase in transgenic tobacco plants reveals control points for flux into phenylpropanoid biosynthesis. *Plant Physiology* 112(4):1617−24
59. Qualley AV, Widhalm JR, Adebesin F, Kish CM, Dudareva N. 2012. Completion of the core β-oxidative pathway of benzoic acid biosynthesis in plants. *Proceedings of the National Academy of Sciences of the United States of America* 109(40):16383−88
60. Lee S, Kaminaga Y, Cooper B, Pichersky E, Chapple C. 2012. Benzoylation and sinapoylation of glucosinolate R-groups in *Arabidopsis*. *The Plant Journal* 72:411−22
61. Klempien A, Kaminaga Y, Qualley A, Nagegowda DA, Widhalm JR, et al. 2012. Contribution of CoA ligases to benzenoid biosynthesis in *Petunia* flowers. *The Plant Cell* 24:2015−30

62. Teotia D, Gaid M, Saini SS, Verma A, Yennamalli RM, et al. 2019. Cinnamate-CoA ligase is involved in biosynthesis of benzoate-derived biphenyl phytoalexin in *Malus* × *domestica* 'Golden Delicious' cell cultures. *The Plant Journal* 100:1176–92
63. Bussell JD, Reichelt M, Wiszniewski AAG, Gershenzon J, Smith SM. 2014. Peroxisomal ATP-binding cassette transporter COMATOSE and the multifunctional protein ABNORMAL INFLORESCENCE MERISTEM are required for the production of benzoylated metabolites in *Arabidopsis* seeds. *Plant Physiology* 164:48–54
64. Van Moerkercke A, Schauvinhold I, Pichersky E, Haring MA, Schuurink RC. 2009. A plant thiolase involved in benzoic acid biosynthesis and volatile benzenoid production. *The Plant Journal* 60:292–302
65. Vogel C, Widmann M, Pohl M, Pleiss J. 2012. A standard numbering scheme for thiamine diphosphate-dependent decarboxylases. *BMC Biochemistry* 13:24
66. Gallage NJ, Hansen EH, Kannangara R, Olsen CE, Motawia MS, et al. 2014. Vanillin formation from ferulic acid in *Vanilla planifolia* is catalysed by a single enzyme. *Nature Communications* 5:4037
67. Liu F, Schnable PS. 2002. Functional specialization of maize mitochondrial aldehyde dehydrogenases. *Plant Physiology* 130:1657–74
68. Long MC, Nagegowda DA, Kaminaga Y, Ho KK, Kish CM, et al. 2009. Involvement of snapdragon benzaldehyde dehydrogenase in benzoic acid biosynthesis. *The Plant Journal* 59:256–65
69. Ibdah M, Pichersky E. 2009. *Arabidopsis Chy*1 null mutants are deficient in benzoic acid–containing glucosinolates in the seeds. *Plant Biology* 11:574–81
70. Alvarez MA. 2014. Plant biotechnology for health: from secondary metabolites to molecular farming, vol. 8. Cham: Springer International Publishing. p. 39–48. https://doi.org/10.1007/978-3-319-05771-2
71. Groves RA, Hagel JM, Zhang Y, Kilpatrick K, Levy A, et al. 2015. Transcriptome profiling of khat (*Catha edulis*) and *Ephedra sinica* reveals gene candidates potentially involved in amphetamine-type alkaloid biosynthesis. *Plos One* 10:e0119701
72. Chang CC, Beevers H. 1968. Biogenesis of oxalate in plant tissues. *Plant Physiology* 43:1821–28
73. Yu L, Jiang J, Zhang C, Jiang L, Ye N, et al. 2010. Glyoxylate rather than ascorbate is an efficient precursor for oxalate biosynthesis in rice. *Journal of Experimental Botany* 61:1625–34
74. Cai X, Ge C, Wang X, Wang S, Wang Q. 2018. Expression analysis of oxalate metabolic pathway genes reveals oxalate regulation patterns in spinach. *Molecules* 23(6):1286
75. Truffault V, Fry SC, Stevens RG, Gautier H. 2017. Ascorbate degradation in tomato leads to accumulation of oxalate, threonate and oxalyl threonate. *Plant Journal* 89:996–1008
76. Parsons HT, Fry SC. 2012. Oxidation of dehydroascorbic acid and 2, 3-diketogulonate under plant apoplastic conditions. *Phytochemistry* 75:41–49
77. Yao J, Pang Y, Qi H, Wan B, Zhao X, et al. 2003. Transgenic tobacco expressing *Pinellia ternata* agglutinin confers enhanced resistance to aphids. *Transgenic Research* 12:715–22
78. Jin S, Zhang X, Daniell H. 2012. *Pinellia ternata* agglutinin expression in chloroplasts confers broad spectrum resistance against aphid, whitefly, lepidopteran insects, bacterial and viral pathogens. *Plant Biotechnology Journal* 10(3):313–27
79. Umer N, Naqvi RZ, Rauf I, Anjum N, Asif M. 2020. Expression of *Pinellia ternata* leaf agglutinin under rolc promoter confers resistance against a phytophagous sap sucking aphid, *Myzus persicae*. *Electronic Journal of Biotechnology* 47:72–82