Fitoterapia 140 (2020) 104412

Contents lists available at ScienceDirect

Fitoterapia

journal homepage: www.elsevier.com/locate/fitote

Transcriptome analysis of *Aconitum carmichaelii* and exploration of the salsolinol biosynthetic pathway

Yuxia Yang[b,1], Ping Hu[b,1], Xianjian Zhou[b,1], Ping Wu[b,1], Xinxin Si[b,c,1], Bo Lu[b,1], Yanxi Zhu[a,2], Yanli Xia[a,*,2]

[a] College of Pharmacy and Biological Engineering of Chengdu University, Chengdu 610106, PR China
[b] Sichuan Provincial Key Laboratory of Quality and Innovation Research of Chinese Materia Medica, Sichuan Academy of Traditional Chinese Medicine Sciences, Chengdu 610041, PR China
[c] Sichuan Agricultural University, Chengdu 611134, PR China

ARTICLE INFO

Keywords:
Aconitum carmichaelii
Alkaloid
C19-diterpenoid alkaloids
Salsolinol
Biosynthetic pathways
Varieties

ABSTRACT

Aconitum carmichaelii has been used in traditional Chinese medicine for treating various diseases for several thousand years. Based on the biosynthetic pathway of some alkaloids such as C19-diterpenoid alkaloids and obvious differences in alkaloid content between leaves of two *A. carmichaelii* varieties has been reported, we performed leaves transcriptome analysis of two *A. carmichaelii* varieties. Besides we characterized the biosynthetic pathway of salsolinol. A total of 56 million raw reads (8.28 G) and 55 million clean reads (8.24 G) were obtained from two varieties (Z175 and R184) leaves transcriptome, respectively, and 176,793 unigenes were annotated.

281 and 843 unigenes are involved in the salsolinol biosynthetic pathway and the formation of C19-diterpenoid alkaloids respectively. And including 34 and 24 unigenes are the differentially expressed genes (DEGs) in the biosynthesis pathway for C19-diterpenoid alkaloids and salsolinol between Z175 and R184 respectively, which were target genes to explore differences in C19-diterpenoid alkaloid and salsolinol biosynthesis in Z175 and R184. Thus genes involved in alkaloid biosynthesis and accumulation differ between varieties leaves. The mechanisms underlying the differences and their relevance require further exploration. The results expand our knowledge of alkaloids biosynthesis in *A. carmichaelii* leaves, and provide a theoretical basis for analysis differences in alkaloids biosynthesis patterns in different varieties.

Abbreviations: TCM, Chinese medicine; DA, diterpene alkaloids; MVA, mevalonate pathway; MEP, methylerythritol pathway; CPS, ent-CPP synthases; KS, ent-kaurene synthases; KOX, kaurene oxidases; TFs, transcription factors; nr, NCBI non-redundant; Nt, NCBI nucleotide sequences; Pfam, protein family; KOG, Clusters of euKaryotic Ortholog Groups; COG, Clusters of Orthologous Groups of proteins; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; DEGs, differential expressed genes; DXS, 1-deoxy-D-xylulose 5-phosphatesynthase; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; ISPD, 2-*C*-methyl-D-erythritol 4-phosphate cytidylyltransferase; ISPE, 4-(cytidine- 50-diphospho)-2-*C*-methyl- D-erythritol kinase; ISPF, 2-C– methyl-D-erythritol2,4 -cyclodiphosphate synthase; ISPG, (E)-4-hydroxy-3-methylbut-2-enyldiphosphate synthase; ISPH, (E)-4- hydroxy-3-methylbut-2-enyl diphosphate reductase; IPPI, isopentenyl diphosphate isomerase; AACT, acetoacetyl-CoA thiolase; HMGR, 3-hydroxy-3- methylglutaryl- CoA reductase; HMGS, 3-hydroxy-3-methylglutaryl- CoA synthase; MVK, mevalonate kinase; PMK, phosphomevalonate kinase; MVDD, mevalonate diphosphate decarboxylase; GGPPS, geranylgeranyl pyrophosphate synthase; CPS, ent-copalyl diphosphate synthase; KS, entkaurene synthases; HK, he xokinase; GPI, glucose-6-phosphate isomerase; PFK, 6-phosphofructokinase; ALDO, fructose-bisphosphate aldolase; GAPHD, glyceraldehyde 3-phosphate dehydrogenase; PGK, phosphoglycerate kinase; PGAM, 2,3-bisphosphoglycerate- dependent phosphoglycerate mutase; ENO, enolase; aroF, phospho-2-dehydro-3- deoxyheptonate aldolase; AROB, 3-dehydroquinate synthase; AROD/AROE, 3-dehydroquinate/shikimate dehydratase; AROK, shikimate kinase; AROA, 3-phosphoshikimate 1-carboxyvinyltransferase; AROC, chorismate synthase; CHMU, chorismate mutase; PAT, bifunctional aspartate/glutamate aminotransferase; TYRC, prephenate dehydrogenase; POX, polyphenol oxidase; TYDC, tyrosine decarboxylase

* Corresponding author.
E-mail address: yanlixiafz@163.com (Y. Xia).
[1] Address: Sichuan Provincial Key Laboratory of Quality and Innovation Research of Chinese Materia Medica, Sichuan Academy of Traditional Chinese Medicine Sciences, No. 51, Fourth Section of Renmin South Road, Wuhou District, Chengdu 610,041,People's Republic of China
[2] Address: College of pharmacy and biological engineering of Chengdu University, No. 2025 Chengluo avenue, Longquanyi district, Chengdu 610,106,People's Republic of China

https://doi.org/10.1016/j.fitote.2019.104412
Received 21 September 2019; Received in revised form 28 October 2019; Accepted 4 November 2019
Available online 05 November 2019
0367-326X/

1. Introduction

The *Aconitum* genus (Ranunculaceae), currently including 211 known species, is described as one of the 'Four Pillars' of ancient herbs used in Chinese medicinal healing, of which 400 are distributed in temperate regions of the northern hemisphere [1,2]. The lateral roots of *Aconitum* have been widely used in traditional Chinese medicine (TCM) to treat various diseases and ailments for several thousand years, including fainting, rheumatic fever, painful joints, gastroenteritis, diarrhea, edema, bronchial asthma, and some endocrinal disorders such as irregular menstruation [3]. Approximately 76 species in the *Aconitum* genus are used for herbal medicines in China [4]. *A. carmichaelii* is one of two species officially recorded as an aconite in the Chinese Pharmacopoeia [5]. This perennial plant grows to a height of 60–150 cm, has dark brown roots, palmately divided dark green glossy leaves and violet flowers with helmet-shaped petals [3]. The primary root of *A. carmichaelii* is known as chuanwu (川乌) in TCM, and the lateral root is called fuzi (附子) [6]. *A. carmichaelii* is mainly grown in Sichuan, Yunnan and Shanxi provinces in China [7]. However, the high toxicity risk and narrow therapeutic range limit large-scale medicinal applications [8].

The pharmacological properties of *A. carmichaelii* are attributed to the accumulation of a diverse range of bioactive C20-, C19– and C18-type diterpene alkaloids (DAs). To date, 75 C19-DAs, 18 C20-DAs and 11 other alkaloids have been identified from the root of *A. carmichaelii* [9,10], but the content of toxic DAs in tubers makes this herb particularly dangerous. > 1500 DAs have been isolated from *Aconitum*, but only lappaconitine, 3-acetylaconitine and bulleyaconitine have been clinically used as analgesics in China [11]. Furthermore, the cardio- and neurotoxicity of these DAs makes them potentially lethal, and the improper use of *Aconitum* in China, India, Japan and elsewhere still results in human poisoning [5]. The toxicity of *Aconitum* DAs arises from their effect on voltage-gated sodium channels, the release of neurotransmitters and changes in receptors, and the promotion of lipid peroxidation and cell apoptosis in the heart, liver and other tissues [10]. The principal toxic and pharmacological ingredients of fuzi are aconitine-type C19-DAs [6,10].

In *Aconitum,* DAs and their derivatives, including C18, C19 and C20 classes, are the principal constituents responsible for biological activity [5]. C19-DAs in processed roots of *Aconitum* following hydrolysis are possibly the most beneficial bioactive constituents of fuzi [12]. The majority of C19-type diterpene alkaloid skeletons, which account for most of the alkaloid content of *A. carmichaelii*, are derived from various substitutions on the aconitine skeleton, leading to 75 different known alkaloids [10]. Mevalonate (MVA) and methylerythritol (MEP) pathways play important roles in the biosynthesis of terpenoids, and biosynthesis of C19-DAs in *A. carmichaelii* begins with MVA or MEP biosynthetic pathways, involving geranylgeranyl pyrophosphate synthase (GGPPS) that produce the precursor of DAs [13]. Furthermore, ent-CPP synthases (CPS), ent-kaurene synthases (KS), kaurene oxidases (KOX), cyclases, aminotransferases, monooxygenases, methyltransferase and BAHD acyltransferases act together to synthesis C19-DAs in fuzi, and transcription factors (TFs) appear to regulate the accumulation of DAs. C19-DAs and C18-DAs may be derived biogenetically from C20-DAs [14,15] via the loss of carbon atom(s) [16].

Salsolinol, an alkaloid involved in the black appearance of overripe bananas [17], is biosynthesized by nonenzymatic Pictet-Spengler condensation, in which a phenylethylamine such as dopamine complexes with acetaldehyde to form an imine, which subsequently cyclize to generate a tetrahydroisoquinoline [18]. In medical science, salsolinol (1-methyl-6,7-dihydroxy-1,2,3,4-tetrahydroisoquinoline) is structurally similar to 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine, which is reported to play a role in the development of Parkinson-like syndrome [19]. Although barely soluble in water [20], salsolinol functions in bradycardiac arrhythmia and decreases blood pressure [21–24]. Current knowledge on salsolinol biosynthesis in *Aconitum* is minimal.

In *A. carmichaelii*, C19-type diterpenes are the most abundant DAs, followed by C20-type and other diterpenes [10]. We cultivated two varieties Z175 and R184 over the past ten years in Sichuan academy of traditional Chinese medicine sciences. High-performance liquid chromatography (HPLC) analysis with an evaporative scattering light detector (ELSD) were carried out on leaves from Z175 and R184 varieties in 2016 and 2017, and we identified C19-diterpenoid alkaloids including benzoylaconine, hypaconitine, neoline, mesaconitine, benzoylmesaconine, fuziline, aconitine, talatizamine, mesaconine, hypaconine, Yunaconitine and aconine, as well as the C20-diterpenoid alkaloid songorine, and one other alkaloid, salsolinol(unpublished data). Besides found that compared with R184, Z175 were only with higher salsolinol, and lower benzoylaconine, hypaconitine, neoline, mesaconitine, benzoylmesaconine, fuziline, aconitine, talatizamine, mesaconine, hypaconine, yunaconitine, aconine and songorine(unpublished data). Analysis of expression patterns of unigenes involved in C19-type alkaloid biosynthesis in different tissues showed that flowers and buds are the most transcriptionally active tissues, while roots and leaves are characterized by a distinct set of transcriptionally active unigenes [13]. It's interesting to searching the C19, C20-type alkaloid and other alkaloid gene expression pattern in leaf and root, and also the relationship between leaf and root. So based on the differences in alkaloid content in different *A. carmichaelii* varieties leaves, especially Z175 with only higher salsolinol, and all lower C19 and C20-diterpenoid alkaloid than R184, and difference expression patterns between tissues, the leaves transcriptome of two varieties were investigated firstly. Then through transcriptome data were acquired from two varieties, the salsolinol biosynthesis pathway-related genes were identified in leaves. Finally, enzymes mediating biosynthesis of the main aconitine C19-DAs were compared.

2. Materials and methods

2.1. Plant materials

A. carmichaelii plants (plant material is two varieties Z175 and R184) were grown in natural conditions at Jiangyou Fuzi GAP cultivation base and were identified by Guangming Shu, a researcher of Sichuan academy of traditional Chinese medicine sciences. And the voucher specimens of this material have been deposited in the herbarium of Sichuan academy of traditional Chinese medicine sciences. The leaves were harvested on ice in the month of September 2016 and snap frozen using liquid nitrogen. Then all samples were taken back indoor and stored at −80 °C for RNA extraction.

2.2. RNA isolation, cDNA library preparation and Illumina sequencing

Total RNA was extracted from *A. carmichaelii* leavess using TRIzol reagent (Ambion) and the RNeasy Plus Mini Kit (No. 74134; Qiagen, Hilden, Germany) following the manufacturer's instructions. RNA quality and integrity was monitored on 1% agarose gels. RNA concentration was measured using Qubit® RNA Assay Kit in Qubit 2.0 Flurometer (Life Technologies, CA, USA). RNA of Z175 and R184 *A. carmichaelii* leavess with three biological duplication were used to construct the cDNA library respectively. cDNA library construction and Illumina sequencing of samples were performed at GenexHealth Corporation (Beijing, China). mRNA samples were purified and fragmented using the NEBNext Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA). Random hexamer primers were used to synthesize the first-strand cDNA, followed by synthesis of the second-strand cDNA using buffer, dNTPs, RNase H, and DNA polymerase I at 16 °C for 1 h. After end repair, A-tailing, and the ligation of adaptors, the products were amplified by PCR and quantified precisely using the Qubit DNA Br Assay Kit (Q10211; Invitrogen, Carlsbad, CA, USA). They were then purified in AMPure XP system (Beckman Coulter, Beverly, USA) to obtain a cDNA library, and library quality was assessed on the Agilent

Bioanalyzer 2100 system(Agilent Technologies, CA, USA). The cDNA library was sequenced on the HiSeq2500 platform in GenexHealth, Ltd., Beijing, China.

2.3. Assembly and functional annotation

All raw reads were processed to remove low-quality and adaptor sequences by Trimmomatic (http://www.usadellab.org/cms/index.php?page=trimmomatic). Transcriptome assembly was accomplished based on the claen reads using Trinity [25] with min_kmer_cov set to 2 by default and all other parameters set default after combined the six leaves clean reads. The largest alternative splicing variants in the Trinity results were called unigenes. The transcription factor (TF) encoding transcripts were identified based on similarity search against Plant Transcription Factor Database (http://planttfdb.cbi.pku.edu.cn). The annotation of unigenes was in seven databases which include NCBI non-redundant protein sequences (Nr), NCBI nucleotide sequences (Nt), Protein family (Pfam), Clusters of Orthologous Groups of proteins/euKaryotic Ortholog Groups (COG/KOG), Swiss-Prot, Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Ontology (GO).For the putative protein sequences, we performed tblastx search against the Nt database (https://www.ncbi.nlm.nih.gov/nuccore) with an e-value cut-off of 1.0E-5. Nr database (http://www.ncbi.nlm.nih.gov/genbank/) and Swiss-Prot (http://www. uniprot.org/) were searched by diamond v0.8.22 [26] with an e-value cut-off of 1.0E-5. The candidates associated with DA biosynthesis were identified using previous reports [13–15,27–31]. And the candidate genes in salsolinol biosynthesis pathway are based on KEGG annotation. Clusters of Orthologous Groups of proteins/euKaryotic Ortholog Groups (COG /KOG) with e-value cut-off of 1.0E-3. Pfam (http://pfam.xfam.org/) were searched by HMMER 3.0 package [32] with e-value = 1.0E-2. BLAST2GO was used to obtain Gene Ontology (GO) annotation of assembled unigenes for describing cellular component, molecular function, and biological process with an e-value cut-off of 1.0E-6 [33]. Unigenes were searched in KEGG database (http://www.genome.jp/kegg/pathway.html)with BLAST at e-value $\leq 1.0E-10$. Expression levels were expressed in terms of FPKM values (fragments per kilobase per million reads) [34], which was calculated by RSEM (RNA-Seq by Expectation-Maximization)(Version: v1.2.6)with default parameters [35]. For the three biological replicates, differential expression analysis of two breed sample (Z175 and R184) was performed using the DESeq R package (1.10.1), and genes with an a log2 fold-change of $> 1$and pdaj < 0.05 found by DESeq were assigned as differentially expressed [36]. GO enrichment analysis provided all GO terms generating from the identified the differentially expressed genes (DEGs). Gene Ontology (GO) enrichment analysis of DEGs(both up-and down-regulated between the two varieties) was mapped to GO terms in the Gene Ontology database (http://www.geneontology.org/) and implemented by the GOseq R packages based Wallenius non-central hypergeometric distribution [37]. We used KOBAS [38] software to test the statistical enrichment of differential expression genes in KEGG pathways.

3. Results

3.1. Transcriptome sequence assembly, annotation and differential gene expression

In total, 56 million raw reads (8.28G) and 55 million clean reads (8.24G) were obtained from six transcriptomes Z175 and R184 varieties in three biological replicates, respectively. In Z175, the percentage of reads with q20 and q30 quality scores was 97.32% and 92.61%, compared with 97.61% and 93.26% for R184 (Fig. 1). After trimming adapters, removing low-quality raw sequences using Trimmomatic (http://www.usadellab.org/cms/index.php?page= trimmomatic), splicing and assembly (using Trinity), we obtained 187 million transcripts from the six leaves transcriptomes of *A. carmichaelii*, with an N50 value of 1309 bp, an average length of 853 bp, and a maximal length of 11,085 bp. And 175 million unigenes were obtained with an N50 value of 1393 bp, an average length of 993 bp, and the same maximal length with transcripts. The six raw read datasets have been deposited in the NCBI SRA database under GenBank accession number SRP190869.

From 108,477 unigenes (61.35%), we obtained matches to entries in the NCBI non-redundant (nr) protein database using BLASTx with an e-value cut-off of 1.0E-5 (Fig. 2a). Meanwhile, 54,936 (31.07%) were aligned to the NCBI nucleotide sequences (Nt) database, 82,039 (46.40%) were aligned to the Swiss-Prot database, 77,612 (43.89%) were aligned to the protein family (Pfam) database, 78,500 (44.4%) were aligned to the GO database, and 27,522 (15.56%) were aligned to the Clusters of euKaryotic Ortholog Groups (KOG) database (Fig. 2a). Interestingly, 1837, 24,380 and 40 unigenes were uniquely annotated to Nt, nr and KOG, respectively (Fig. 2c). Furthermore, 176,793 unigenes were annotated to at least one of the searched databases. And 35.4% of unigenes were successfully aligned to *Nelumbo nucifera* genes using BLASTx in Nr database, followed by *Vitis vinifera* (12.5%), *Theobroma cacao* (2.9%), *Jatropha curcas* (2.4%) and *Citrus sinensis* (2.0%), and 44.8% were aligned with other species (Fig. 2b).

We used GO annotations to classify the 78,500 unigenes into functional groups using BLAST2GO where *p*-values calculated by hypergeometric distribution tests and e-values were $< 1.0E-5$. In the *A. carmichaelii* transcriptome data, biological process accounted for the majority of GO annotations (47.33%, 179,968 genes), followed by cellular component (28.10%, 106,856 genes) and molecular functions (24.57%, 93,447 genes). In the biological process category, GO terms related to cellular process (23.43%) and metabolic process (22.29%) were most abundant, followed by single-organism process (16.62%), biological regulation (7.14%) and regulation of biological process (6.61%). In the cellular component category, cell (19.88%) and cell part (19.87%) were dominant, followed by organelle (13.39%), macromolecular complex (12.69%) and membrane (10.48%). In the molecular function category, binding (48.70%) and catalytic activity (38.03%) were the most represented, followed by catalytic activity and transporter activity (4.67%), structural molecule activity (2.50%) and nucleic acid binding transcription factor activity (2.11%; Fig. 3a).

A total of 30,657 transcripts were classified into 26 KOG categories. General function prediction (7602 transcripts, 24.80%) was the dominant KOG category, followed by posttranslational modification, protein turnover and chaperones (4267 transcripts, 13.91%) and general function prediction only (3335 transcripts, 10.88%; Fig. 3b).

To elucidate active biosynthesis pathways in *A. carmichaelii*, annotation of nr data using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database identified 43,695 genes assigned to five main categories representing 130 biological pathways. The highest number of KO identifiers were linked to metabolism (18,951 genes) followed by genetic information processing (9809 genes), cellular processes (1813 genes), organismal systems (1410 genes) and environmental information processing (1372 genes). Pathways with the largest number of KO identifiers were: translation (4070); carbohydrate metabolism (3808); folding, sorting and degradation (3133); overview (2895); amino acid metabolism (2210) and lipid metabolism (1974; Fig. 3c).

A total of 4866 transcripts (2.75%) were assigned to 80 TF families. Among these, MYB (308), C3H (249), AP2-EREBP (199), NAC (195), Orphan (190) and C2H2 (185) were the most abundant. Furthermore, 322 (6.62%) differentially expressed genes (DEGs) encoding TFs potentially involved in the regulation of various physiological processes and biochemical pathways in *A. carmichaelii* were identified (Supplemental Table S1).

Expression of read-mapped genes was analysed based on the fragments per kilobase of transcript per million mapped reads (FPKM) value for each clean read. Transcripts with $|\log2FoldChange| > 1$ and padj < 0.05 were considered DEGs. Among the 176,793 identified unigenes, ~6873 DEGs were detected, of which 3470 were down-

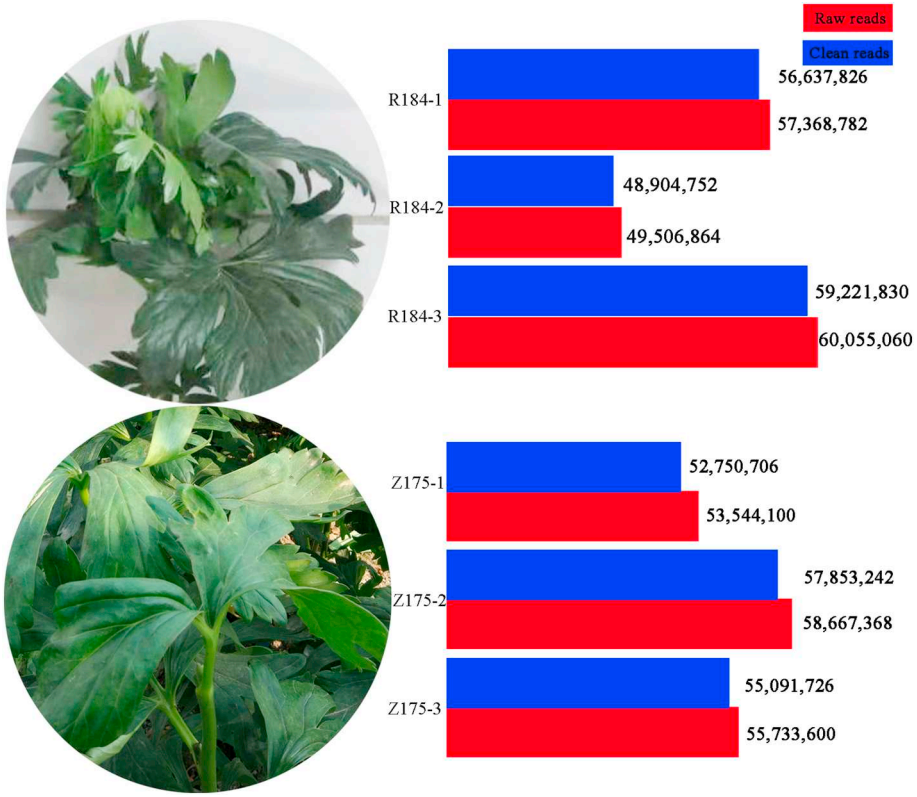


Fig. 1. Raw and clean reads obtained from transcriptome analysis of different *A. carmichaelii* tissues.

regulated and 3403 were upregulated in the two *A. carmichaelii* varieties. A total of 4175 DEGs were linked to 356 metabolic pathways (p-value < .05). Based on DEGs between Z175 and R184 varieties, GO analysis showed that 2076 GO accessions matched biological process, 454 matched cellular component, and 993 matched molecular function categories (Fig. 4). KEGG analysis identified 320 pathways related to 4175 annotated DEGs (p-value < .05; Table S2).

3.2. Characterization and expression analysis of unigenes involved in C19-type diterpene biosynthesis in Z175 and R184

Based on the results of previous studies [13,27], putative unigenes involved in the aconitine biosynthesis pathway were identified, including 49 unigenes encoding 1-deoxy-D-xylulose-5-phosphatesynthase (DXS), two of which were more highly expressed in Z175 than R184, 1 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR), 10 2-*C*-methyl-D-erythritol 4-phosphate cytidylyltransferase (ISPD), five 4-(cytidine-50-diphospho)-2-*C*-methyl-D-erythritol kinase (ISPE), two 2-*C*-methyl-D -erythritol2,4-cyclodiphosphate synthase (ISPF), 1I (E)-4-hydroxy-3-methylbut- 2-enyldiphosphate synthase (ISPG), two (E)-4-hydroxy-3-methylbut-2- enyldiphosphate reductase (ISPH), and four isopentenyl diphosphate isomerase (IPPI) in the MEP pathway (Table S3). Furthermore, 14 acetoacetyl-CoA thiolase (AACT) and seven 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR) involved in the MVA pathway were identified, including two more highly expressed in Z175, along with two 3-hydroxy-3-methylglutaryl-CoA synthase, (HMGS), three mevalonate kinase (MVK), six phosphomevalonate kinase (PMK), and two mevalonate diphosphate decarboxylase (MVDD; Table S4). Five geranylgeranyl pyrophosphate synthase (GGPPS), seven CPS, 24 KS (Cluster-219.18055 more highly expressed in Z175), 21 KOX, of which four are more highly expressed in Z175, 84 cyclases, of which Cluster-219.66822 and Cluster-219.62280 were higher expressed than Z175, and Cluster-219.80709 and Cluster-219.73820 higher than R184, 131 aminotransferases, of which Cluster-219.46073 was higher expressed than R184, and Cluster-219.51635 higher than Z175, 190 monooxygenases, of which Cluster-219.114355, Cluster-219.51666, Cluster-219.28553 were higher expressed in Z175 than R184, and Cluster-219.140395, Cluster-219.35262, Cluster-219.119209, Cluster-219.2681, Cluster-26,892.0, Cluster-219.131122 higher than Z175, 73 BAHD acyltransferases of which Cluster-219.50148, Cluster-219.20417, Cluster-219.64576, Cluster-219.19319, Cluster-219.90737, Cluster-20,871.1, Cluster-219.44720, Cluster-219.48909, Cluster-219.4883 were higher expressed in Z175 than R184, and Cluster-219.10123, Cluster-219.138567, Cluster-219.32670, Cluster-219.138566, Cluster-219.99337 higher than Z175, and 200 methyltransferases, of which Cluster-219.14882, Cluster-219.17602 were higher expressed in R184 than Z175, and Cluster-22,914.0 higher than R184, were also identified (Table S5). Expression of all the above higher or lower unigenes was statistically significant (log2 fold-change > 1.0 and pdaj < 0.05), and the sequence of them was in Supplemental file 1. Expression levels of unigenes involved in the aconitine biosynthesis pathway in the two varieties are shown in Table S3-S5. Expression levels of unigenes in the isopentenyl diphosphate pathway were highest, followed by MEP and MVA pathways (Table S3-S5).

3.3. Candidate genes encoding enzymes involved in the salsolinol biosynthesis pathway

Salsolinol is another important aconitine in *A. carmichaelii*. A schematic diagram of the proposed pathway for the biosynthesis of salsolinol is shown in Fig. 5. Precursors for the biosynthesis of salsolinol are derived from D-glucose through glycolysis that generates phosphoenolpyruvate, then through the shikimate pathway that synthesizes chorismate, then through dopamine, finally yielding salsolinol. In more detail, the glycolysis pathway includes 19 hexokinase (HK) with Cluster-219.117566 higher expressed in Z175, 13 glucose-6-phosphate isomerase (GPI) with Cluster-219.75274 higher expression in Z175 than R184, 30 6-phosphofructokinase (PFK) 31 fructose-bisphosphate

Protein Databases	Number of Unigenes	Percentage (%)
nr	108477	61.35
NT	54936	31.07
SwissProt	82039	46.4
PFAM	77612	43.89
GO	78500	44.4
KOG	27522	15.56
Total	176793	100

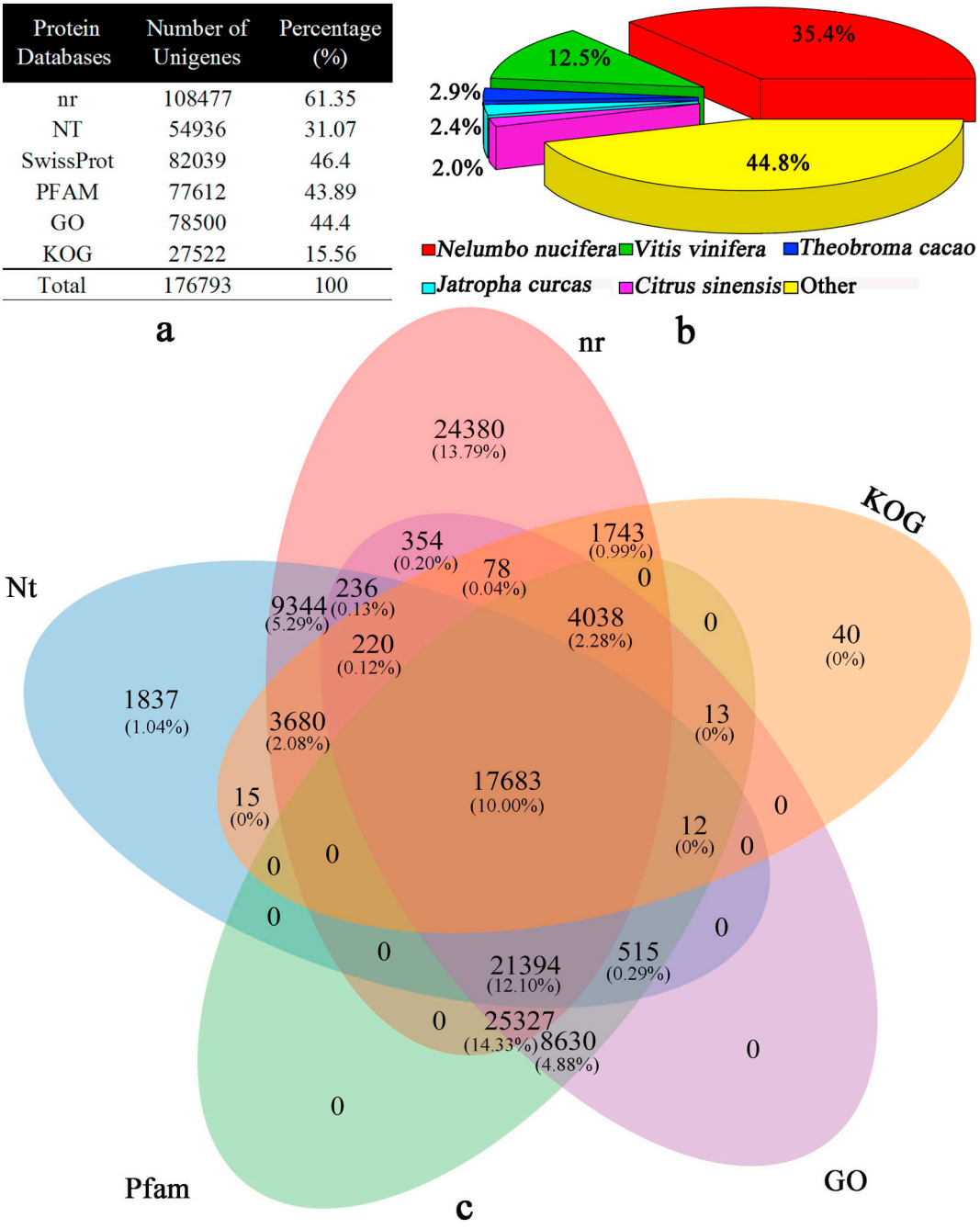


Fig. 2. Statistical overview of functional annotation. (a) Functional annotation of unigenes using various public protein databases. (b) Unigenes in known species identified by BLASTx searches. (c) Venn diagram showing functional annotation of unigenes using NCBI non-redundant protein sequences (Nr), NCBI nucleotide sequences (Nt), Swiss-Prot, Protein family (Pfam), Clusters of euKaryotic Ortholog Groups (KOG) and Gene Ontology (GO) databases.

aldolase (ALDO), of which Cluster-219.73817, and Cluster-28,961.0 were higher expression in Z175 than R184, and Cluster-219.98335, Cluster-219.14890, Cluster-219.9488 and Cluster-219.71706 higher expressed in R184, nine glyceraldehyde-3-phosphate dehydrogenase (GAPHD), 32 phosphoglycerate kinase (PGK) with Cluster-219.80323 and Cluster-219.80326 higher expression in Z175, one 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase (PGAM) and 23 enolase (ENO) enzymes (Table S6). In the shikimate pathway, there are 22 phospho-2-dehydro-3-deoxyheptonate aldolase (AROF) enzymes, four of which, Cluster-219.50402, Cluster-219.4979, Cluster-219.95746, Cluster-219.78492 and Cluster-219.78493, were more highly expressed in Z175, two 3-dehydroquinate synthase (AROB), nine 3-dehydroquinate/shikimate dehydratase (AROD/AROE), 13 shikimate kinase (AROK), of which Cluster-219.54689 and Cluster-219.77462 were more highly expressed in R184 and Z175, two 3-phosphoshikimate 1-carboxyvinyltransferase (AROA), and 27 chorismate synthase (AROC; Table 1). The chorismate pathway includes seven chorismate mutase (CHMU), five bifunctional aspartate/glutamate aminotransferase (PAT), seven prephenate dehydrogenase (TYRC), 23 polyphenol oxidase (POX), including seven unigenes, Cluster-22607.0, Cluster-219.82880, Cluster-219.54918, Cluster-219.54917, Cluster-219.50283, Cluster-17242.0 and Cluster-219.92608 more highly expression in Z175 than R184, and six tyrosine decarboxylase (TYDC) enzymes (Table 2). All sequences of expressed statistically significant unigenes between Z175 and R184 were in Supplemental file 2. Comparison of unigene expression levels in glycolysis, shikimate and chorismate pathways revealed highest expression levels in glycolysis, followed by shikimate and chorismate pathways.

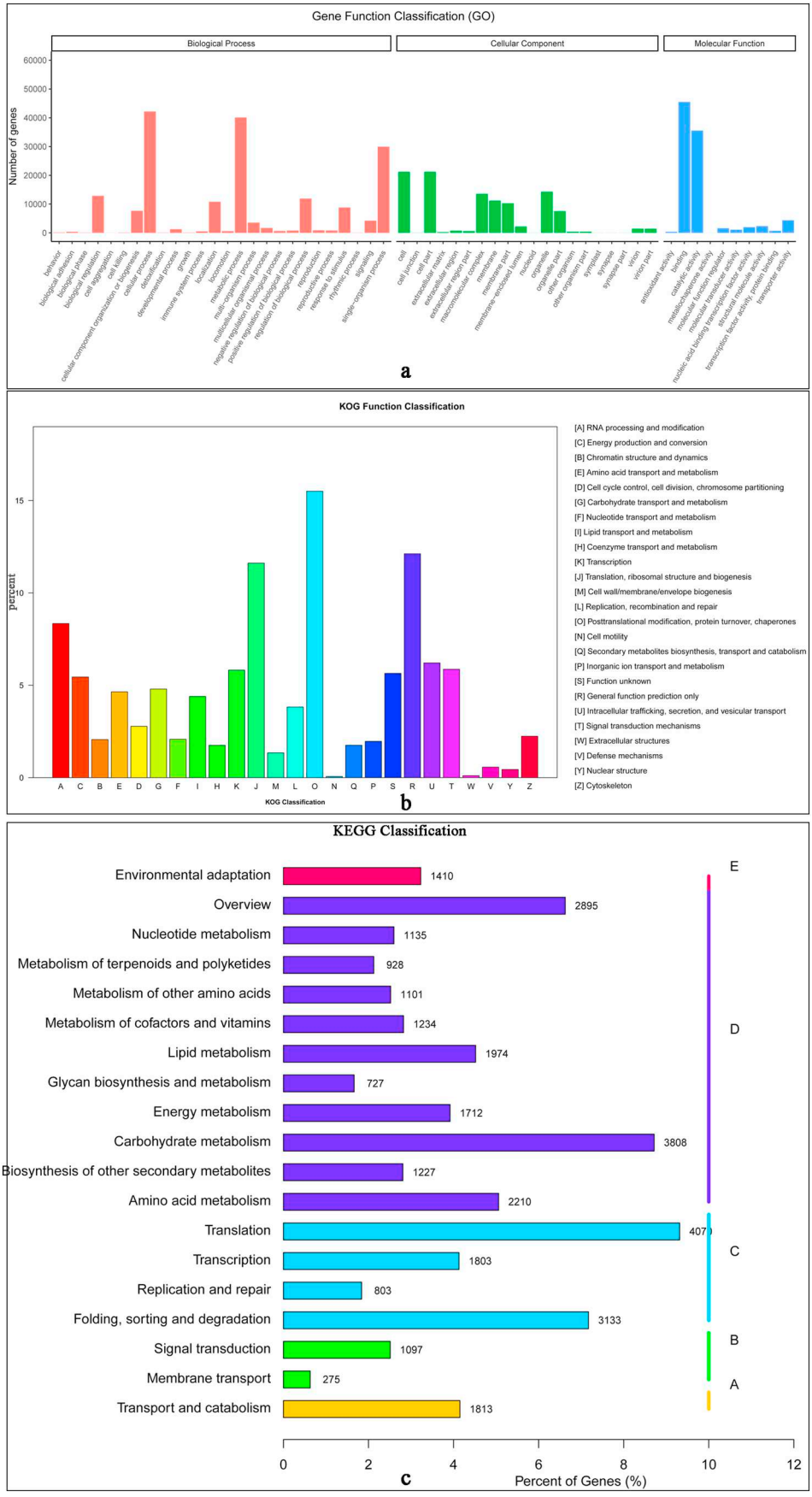


Fig. 3. GO, KOG and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional classification results. (a) GO function classification. (b) KOG functional classification. (c) KEGG functional classification.

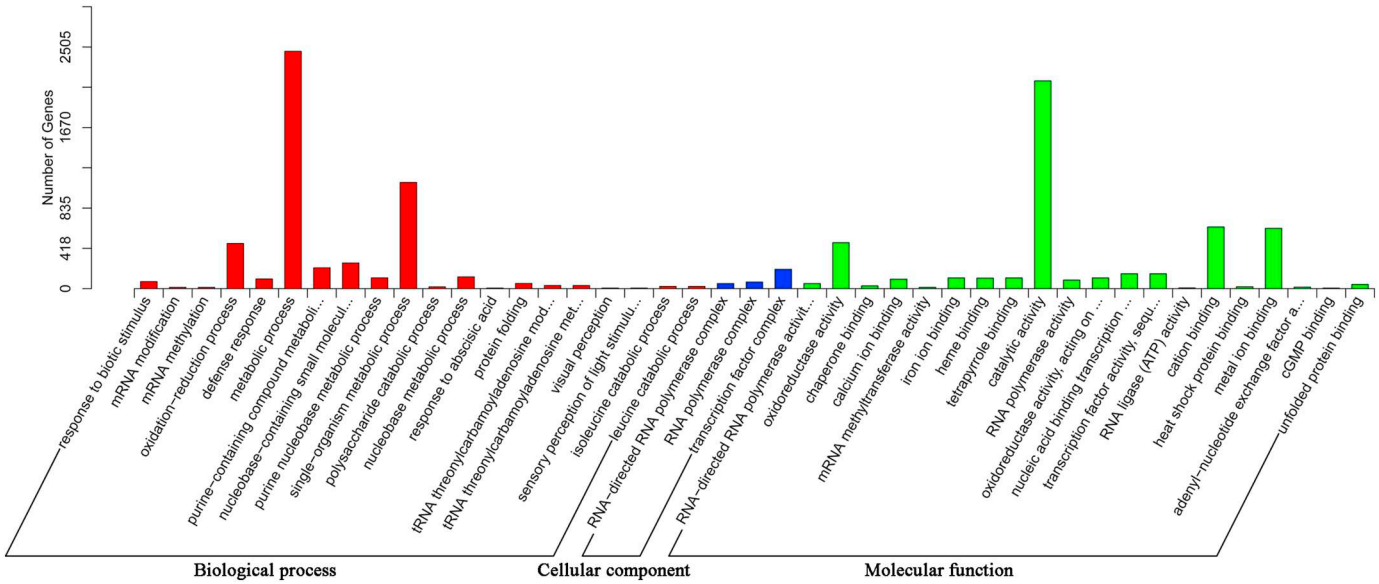


Fig. 4. GO enrichment analysis of DEGs between Z175 and R184 varieties.

Salsolinol was detected in both Z175 and R184, and at a higher concentration in Z175. Salsolinol biosynthesis in *A. carmichaelii* is derived from the biosynthesis of alkaloids generated by glycolysis (map00010), the shikimate pathway, phenylalanine, tyrosine, and tryptophan biosynthesis (map00400) and isoquinoline alkaloid biosynthesis (map00950). In the salsolinol biosynthesis pathway, genes encoding enzymes related to glycolysis were the most highly expressed, followed by shikimate and chorismate pathway genes.

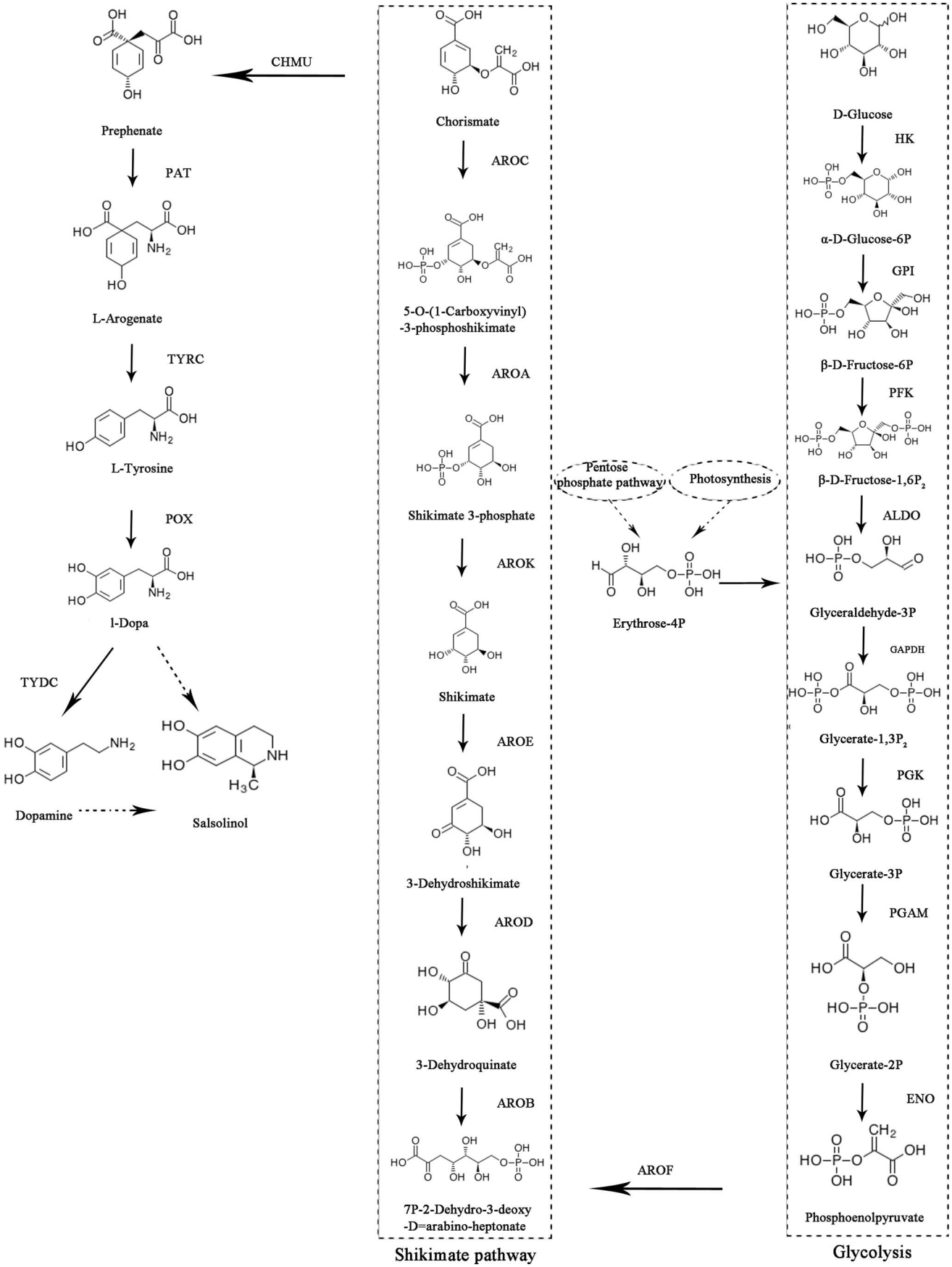
The glycolysis pathway and its enzymes convert glucose to pyruvic acid using the oxidative potential of NAD^+, and this pathway is among the most ancient molecular metabolic networks [39]. In glycolysis, glucose is converted to phosphoenolpyruvate (PEP) by eight central glycolysis enzymes. This was the most well-represented biological process, with one PGAM, 19 HK, 13 GPI, 30 PFK, 31 ALDO, nine GAPHD, 32 PGK and 23 ENO enzymes, and six unigenes were highly expressed in Z175, along with four in R184, while 7/31 ALDO, 6/32 PGK, 2/23 ENO, 1/19 HK, 2/13 GPI, 1/9 GAPHD and 1/30 PFK genes were among the top 20 most highly expressed enzymes related to glycolysis. Most previous studies related to the biosynthesis of C19-diterpenoid alkaloids indicated changes in the abundance of enzymes involved in carbohydrate metabolism in plants under abiotic stress. Specifically, expression of enzymes catalyzing steps in glycolysis such as GAPDH, PGK and PGM, and especially ENO are upregulated [40]. In the case of ENO, this occurs in both the cytoplasm, where it functions as a key glycolytic enzyme, and in plastids which have their own glycolytic pathway [41]. Glycolytic enzymes have acquired additional non-glycolytic functions, as exemplified by HK, GAPHD and ENO, which function in transcriptional regulation, while glucose-6-phosphate isomerase (GPI) functions in the stimulation of cell motility, and HK and GAPHD regulate apoptosis [42].

4. Discussion

Advantages such as high-throughput, accuracy and reproducibility have resulted in transcriptome sequencing becoming established as a powerful technology [43], now widely employed for deep sequencing and exploring the molecular mechanisms of various biological phenomena. Two key studies involving *A. carmichaelii* transcriptome analysis have been reported, one exploring expression in flower, bud, leaves and root tissues [13], and the other analysing rootstock and leaves tissues [27]. Both studies investigated C19– and C20-diterpenoid alkaloid biosynthetic pathways. In the study, we obtained leaves transcriptome data from two varieties of *A. carmichaelii* (Z175 and R184) to investigate potential differences in alkaloid biosynthesis and probe the physiological significance. Leaves transcriptome data showed that the average length of unigenes was 993 bp, considerably longer than 595 bp [27] and 642 bp [13] reported previously. From the six leaves transcriptome datasets collected in triplicate in the present work, 108,477 transcripts were annotated, compared with 128,183 transcripts from transcriptome data from four tissues [13] and 95,812 transcripts from two tissues [27]. The comparative results show that our present transcriptome sequencing and annotation analysis information is more abundant than these previous studies. According to the latest plant TF categories [44], 4866 transcripts were assigned to 80 TF families, compared with only 644 transcripts assigned to 39 branches in previous work [27], but IF, MYB and C3H TFs were essentially the same. In medicinal plants, MYB and C3H TFs regulate specialized metabolic pathways, as do WRKY, ERF and bHLH TFs [13,45,46].

A. carmichaelii is a medicinal plant of high therapeutic value due to the presence of alkaloids, among which C19– and C20-diterpenoid alkaloids are the predominant groups. The active compounds in *A. carmichaelii* endow broad biological activities, including effects on the cardiovascular system, anti-inflammation and analgesic action, anti-tumor activity, immune system effects, hypoglycemic and hypolipidemic effects, anti-aging, kidney protection and energy metabolism [10]. In the biosynthesis of aconitine, both MVA and MEP biosynthetic pathways contribute to the biosynthesis of DAs in plants [47]. A total of 843 unigenes are involved in the biosynthesis of C19-diterpenoid alkaloids, including 22 unigenes associated with the MEP pathway and 34 unigenes related to the MVA pathway, considerably more than the number identified in a previous four-tissue study [13].

In the aconitine biosynthesis pathway, unigenes related to the isopentenyl diphosphate pathway were the most highly expressed in *A. carmichaelii* leaves, followed by those in the MEP and MVA pathways, similar to previous results [13]. Several studies have reported the tissue compartmentalization of precursors, intermediates and final products in highly specialized cells, as well as in different plant tissues [48,49], and *A. carmichaelii* may adopt a similar strategy for synthesizing aconitine-type DAs, which may explain the observed differences in the expression patterns of unigenes involved in different pathways. In the MEP pathway, DXS is believed to act as a bottleneck enzyme [50] and 12 unigenes representing 60 DXS enzymes were among the top 20 most highly expressed according to our current results, which indicates the importance of MEP pathways in *A. carmichaelii* leaves. The MVA



(caption on next page)

Fig. 5. Salsolinol proposed biosynthesis pathway. Glycolysis pathway: hexokinase: HK (EC: 2.7.1.1); glucose-6-phosphate isomerase: GPI(EC: 5.3.1.9); 6-phosphofructokinase: PFK(EC: 2.7.1.11); fructose-bisphosphate aldolase: ALDO(EC: 4.1.2.13); glyceraldehyde 3-phosphate dehydrogenase: GAPHD(EC: 1.2.1.9); phosphoglycerate kinase:PGK(EC: 2.7.2.3); 2,3-bisphosphoglycerate-dependent phosphoglycerate mutase; PGAM(EC: 2.7.2.3); ENO: enolase(EC:4.2.1.11); phospho-2-dehydro-3- deoxyheptonate aldolase: aroF(EC: 2.5.1.54); Shikimate pathway: 3-dehydroquinate synthase: AROB(EC: 4.2.3.4); 3-dehydroquinate/shikimate dehydratase:AROD(EC: 4.2.1.10)/AROE(EC: 4.2.1.118); shikimate kinase: AROK(EC: 2.7.1.71) 3-phosphoshikimate 1-carboxyvinyltransferase: AROA (EC: 2.5.1.19); chorismate synthase: AROC (EC: 4.2.3.5); CHMU: chorismate mutase(EC: 5.4.99.5); bifunctional aspartate aminotransferase and glutamate: PAT(EC: 2.6.1.1); prephenate dehydrogenase: TYRC(EC: 1.3.1.12); polyphenol oxidase:POX(EC: 1.10.3.1); tyrosine decarboxylase: TYDC(EC: 4.1.1.25).

Table 1
Unigenes involved in shikimat pathway of salsolinol biosynthesis pathway in *A. carmichaelii*.

Pathway	Enzyme	Enzyme name	Unigene	Unigene length	NR Evalue	Z175 readcount	R184 readcount	Log2 Fold Change	Padj	Significant
Shikimate	AROF (22)	phospho-2-dehydro-3-deoxyheptonate aldolase	Cluster-219.70422	1416	2.20E-189	987.779	586.544	0.754	0.129	FALSE
			Cluster-219.50402	997	3.50E-138	24.846	0.000	6.989	0.000	TRUE
			Cluster-219.76376	1448	7.50E-222	1094.075	642.408	0.764	0.595	FALSE
			Cluster-219.4979	682	1.50E-68	30.730	0.000	7.293	0.000	TRUE
			Cluster-219.85919	905	1.10E-24	8.002	2.799	1.479	0.986	FALSE
			Cluster-219.76377	1065	1.80E-23	40.305	49.895	−0.278	0.986	FALSE
			Cluster-219.95746	1596	3.20E-189	136.163	50.428	1.393	0.034	TRUE
			Cluster-219.66942	2214	2.30E-177	1083.541	702.647	0.627	0.170	FALSE
			Cluster-219.68452	2244	2.30E-177	290.738	147.412	0.996	0.126	FALSE
			Cluster-219.90735	795	1.50E-91	20.197	11.820	0.861	0.898	FALSE
			Cluster-219.90734	777	6.60E-31	3.628	2.678	0.571	0.986	FALSE
			Cluster-219.90736	514	1.50E-84	14.424	12.661	0.173	0.986	FALSE
			Cluster-219.90731	373	4.80E-56	0.364	2.484	−2.974	0.771	FALSE
			Cluster-219.90733	410	7.90E-68	7.943	4.459	0.840	0.986	FALSE
			Cluster-219.78492	1449	1.10E-143	148.295	6.287	4.434	0.000	TRUE
			Cluster-219.78493	1123	6.30E-152	442.821	7.578	5.910	0.000	TRUE
			Cluster-219.55023	1625	1.20E-143	12.687	5.843	1.132	0.901	FALSE
			Cluster-219.76378	1648	1.20E-143	5.741	7.624	−0.418	0.986	FALSE
			Cluster-219.76379	940	1.10E-24	35.417	8.814	1.964	0.064	FALSE
			Cluster-219.76371	1474	1.60E-219	932.282	286.674	1.695	0.069	FALSE
			Cluster-219.76375	1163	7.00E-170	532.506	395.147	0.425	0.986	FALSE
			Cluster-219.78348	481	1.60E-19	2.304	2.941	−0.142	0.993	FALSE
	AROB (2)	3-dehydroquinate synthase	Cluster-219.68567	1576	1.00E-171	293.621	297.835	−0.027	0.995	FALSE
			Cluster-219.58881	1586	1.00E-171	1178.676	1136.651	0.050	0.986	FALSE
	AROD/ AROE (9)	3-dehydroquinate/ shikimate dehydratase	Cluster-219.31617	622	1.70E-37	20.136	16.458	0.230	0.986	FALSE
			Cluster-219.98650	631	2.40E-39	26.939	32.203	−0.295	0.986	FALSE
			Cluster-219.117612	855	2.20E-112	135.940	171.138	−0.347	0.976	FALSE
			Cluster-219.115502	634	8.80E-18	3.661	13.785	−1.901	0.727	FALSE
			Cluster-219.111892	828	1.80E-103	67.842	90.047	−0.424	0.986	FALSE
			Cluster-219.44644	1785	1.10E-185	80.404	57.276	0.494	0.975	FALSE
			Cluster-219.53108	1767	1.10E-185	366.201	561.737	−0.623	0.595	FALSE
			Cluster-219.53107	628	1.30E-37	22.827	39.310	−0.813	0.904	FALSE
			Cluster-219.120274	628	3.40E-38	7.130	3.130	1.106	0.986	FALSE
	AROK (13)	shikimate kinase	Cluster-219.96711	548	4.30E-21	23.179	46.764	−0.978	0.631	FALSE
			Cluster-219.75068	1254	8.20E-92	533.148	439.217	0.278	0.956	FALSE
			Cluster-219.48198	929	5.00E-70	3.093	2.221	0.337	0.986	FALSE
			Cluster-219.48197	1043	1.90E-86	39.457	32.860	0.275	0.986	FALSE
			Cluster-219.49998	585	2.60E-16	5.737	7.671	−0.469	0.986	FALSE
			Cluster-219.54689	1327	1.10E-91	0.000	17.765	−6.746	0.001	TRUE
			Cluster-219.77462	1312	8.60E-92	409.405	70.477	2.560	0.000	TRUE
			Cluster-219.77463	492	1.50E-17	21.660	22.493	−0.087	0.987	FALSE
			Cluster-219.42507	541	5.40E-16	1.091	1.721	−0.698	0.986	FALSE
			Cluster-219.110751	537	2.50E-21	50.560	29.478	0.728	0.810	FALSE
			Cluster-219.106859	1473	2.90E-112	364.712	438.334	−0.271	0.949	FALSE
			Cluster-219.50000	666	7.70E-12	20.422	22.756	−0.215	0.986	FALSE
			Cluster-219.77461	1269	1.10E-91	3.039	151.252	−5.580	0.000	TRUE
	AROA (2)	3-phosphoshikimate 1-carboxyvinyltransferase	Cluster-219.53999	2137	1.40E-232	580.358	427.825	0.435	0.788	FALSE
			Cluster-219.54000	2136	1.40E-232	74.586	0.000	8.573	0.292	FALSE
	AROC (27)	chorismate synthase	Cluster-219.1696	1747	3.00E-217	3.175	0.482	2.353	0.859	FALSE
			Cluster-4218.0	1296	4.10E-54	3.812	2.785	0.373	0.986	FALSE
			Cluster-219.70284	1401	1.50E-20	4.217	7.898	−0.963	0.960	FALSE
			Cluster-18,528.0	1157	1.80E-149	0.817	0.509	0.389	0.991	FALSE
			Cluster-219.72560	1866	3.80E-234	0.000	1.754	−3.462	0.960	FALSE
			Cluster-219.2484	2287	5.20E-201	25.197	27.189	−0.129	0.993	FALSE
			Cluster-16,761.0	1204	4.70E-60	1.295	0.722	0.661	0.986	FALSE
			Cluster-219.127477	2389	3.10E-180	8.193	6.128	0.363	0.986	FALSE
			Cluster-219.127478	2309	3.00E-180	20.332	10.144	0.947	0.986	FALSE
			Cluster-219.127479	2221	2.70E-138	52.510	60.627	−0.208	0.986	FALSE
			Cluster-219.98159	1910	2.50E-140	12.405	13.315	−0.119	0.986	FALSE
			Cluster-219.34273	1153	3.60E-110	42.996	46.746	−0.100	0.986	FALSE
			Cluster-219.37391	1526	2.00E-172	77.856	51.714	0.562	0.966	FALSE
			Cluster-219.108691	498	3.80E-08	9.540	8.087	0.263	0.986	FALSE

(*continued on next page*)

Table 1 (*continued*)

Pathway	Enzyme	Enzyme name	Unigene	Unigene length	NR Evalue	Z175 readcount	R184 readcount	Log2 Fold Change	Padj	Significant
			Cluster-219.127475	2207	5.00E-201	65.792	97.066	−0.561	0.954	FALSE
			Cluster-219.139826	1230	8.00E-116	14.052	18.452	−0.442	0.986	FALSE
			Cluster-219.103461	1164	3.60E-158	64.412	90.052	−0.508	0.881	FALSE
			Cluster-219.103460	1412	1.20E-131	2.786	13.612	−2.249	0.739	FALSE
			Cluster-219.56676	1592	4.60E-196	283.565	237.030	0.249	0.986	FALSE
			Cluster-219.92842	1708	3.60E-183	413.452	256.299	0.685	0.161	FALSE
			Cluster-219.130746	1237	3.60E-116	4.392	16.128	−1.879	0.890	FALSE
			Cluster-219.103459	1380	1.20E-131	16.792	28.999	−0.744	0.975	FALSE
			Cluster-219.103456	1567	1.00E-155	29.700	40.782	−0.487	0.986	FALSE
			Cluster-219.103457	1534	1.30E-131	8.712	24.885	−1.488	0.416	FALSE
			Cluster-219.103454	320	4.20E-24	1.455	0.254	1.890	0.986	FALSE
			Cluster-18,286.0	1184	4.40E-111	4.167	6.664	−0.706	0.986	FALSE
			Cluster-12,929.0	1124	2.60E-60	0.364	0.763	−1.422	0.986	FALSE

pathway contributes metabolic intermediates for the biosynthesis of sesquiterpenoids including aromatic compounds, and plays an important role in attracting pollinators and seed disseminators [51,52]. HMGR is believed to be a bottleneck enzyme in MVA pathways [50], and four unigenes representing eight HMGR enzymes were among the top 20 most highly expressed in our current results, which indicates the importance of MVA pathways.

In the isopentenyl diphosphate pathway, seven aminotransferases, seven cyclases, two methyltransferase, two BAHD acyltransferases, one kaurene synthase and one ent-kaurene oxidase were among the top 20 most highly expressed, but the significance for regulating the biosynthesis of aconitine remains unknown. Additionally, unigenes annotated as GGPPS and CDPS were highly expressed in *A. carmichaelii* leaves tissue [13], consistent with GGPPS but not CDPS in the present work. Several studies report that CDPS is involved in tissue-specific accumulation of DAs [53,54], but the reason for the low expression of CDPS observed in the present work needs to be further explored.

The shikimate pathway is the key process by which aromatic amino acids are synthesized, and it is also important for salsolinol biosynthesis. Aromatic amino acids are synthesized via the shikimate pathway followed by the branched aromatic amino acid metabolic pathway, with chorismate serving as a major branch point intermediate metabolite [55]. The conversion of PEP and erythrose-4P to chorismate comprises seven reactions catalyzed by six enzymes in the shikimate pathway. Herein, 22 AROF, two AROB, nine AROD/AROE, 13 AROK, two AROA and 27 AROC enzymes were linked to the shikimate pathway, with five unigenes more highly expressed in Z175 and one more highly expressed in R184. All AROB, 9/22 AROF, 1/2 AROA, 3/27 AROC, 3/13 AROK and 2/9 AROD/AROE were in the top 20 most highly expressed among all enzymes in the shikimate pathway. AROF activity may be central to the ability of the shikimate pathway to compete for PEP and erythrose-4P with glycolysis and the non-oxidative pentose phosphate pathway [55]. Thus, high expression of aroF may be conducive to successful salsolinol biosynthesis. Indeed, aroF genes were abundant among the identified unigenes (22), and all four unigenes in Z175 were highly expressed, consistent with higher salsolinol content in Z175 than R184, and any of these four DEGs can be selected as target genes. AROB, the second enzyme in this pathway, converts 3-deoxy-d-arabino-heptulosonate-7-phosphate into 3-dehydroquinate, both of which are highly expressed in *A. carmichaelii.* The actions of AROD/AROE lead to shikimate, bifunctional activity has been characterized in *Solanum lycopersicum* [56], and the crystal structure of the *Arabidopsis* enzyme with shikimate bound at the AROE site and tartrate at the AROD site has recently been elucidated [57]. The enzyme encoded by AROK converts shikimate to shikimate 3-phosphate. Interestingly, *Oryza sativa* AROK genes are differentially expressed in specific developmental stages and in response to biotic stress [58], and plant AROK activity is sensitive to changes in cellular ATP, suggesting that plant AROK acts as a regulatory factor in the shikimate pathway, facilitating metabolic flux toward specific pools of secondary metabolites [55]. The activity of AROA leads to the synthesis of enolpyruvylshikimate-3-phosphate (EPSP), and AROC converts EPSP to chorismate.

In chorismate pathway, there are seven CHMU, five PAT, seven TYRC, 23 POX and six TYDC enzymes that complete the biosynthesis route of salsolinol(Fig. 5) of which 14/23 POX, 2/5 PAT, 1/7 CHMU and 1/6 TYDC enzymes were in the top 20 most highly expressed. The major biosynthetic route for phenylalanine in plants proceeds via arogenate [55]. CHMU converts chorismate to prephenate, and prephenate is then transaminated to arogenate, which is subsequently converted to tyrosine by TYRC [59,60]. Tyrosine is synthesized de novo in plants via two alternative routes mediated by a TyrA family enzyme and prephenate in the shikimate pathway [61], and intermediates serve as precursors of a wide range of secondary metabolites [59]. The tyrosinase activity of POX is believed to be the key factor mediating the initial L-DOPA conversion [62]. Many studies demonstrate a correlation between POX and metabolism, such as betalain pigment formation [63]. In *A. carmichaelii,* POX enzymes are abundant (23), and all seven unigenes are highly expressed in Z175, consistent with the higher salsolinol content in Z175 than R184, and all seven DEGs can be selected as target genes. Dopamine is derived from L-Dopa by tyrosine decarboxylase [64].

Interestingly, coryneine chloride, which is functionally equivalent as salsolinol [22], was isolated from *A. carmichaelii* in 1979, and the tethrahydroisoquinoline salsolinol is present in banana and biosynthesized from dopamine [65], further rationalizing salsolinol biosynthesis in *A. carmichaelii.* However, it remains unclear how dopamine biosynthesis is linked to salsolinol in *A. carmichaelii*. In the mammalian brain, salsolinol can be enzymatically synthesized by salsolinol synthase from dopamine and acetaldehyde, but direct evidence of its biosynthesis is still lacking [66]. Synthesis can occur nonenzymatically as mentioned above. Thus, biosynthesis of salsolinol is not completely understood, although it is known to be derived from dopamine or L-Dopa.

5. Conclusion

A total of 281 unigenes were identified as potentially involved in the salsolinol biosynthetic pathway in *A. carmichaelii*, with 24 DEGs between Z175 and R184 varieties, including 158 in glycolysis and 75 in the shikimate pathway. A total of 843 candidate unigenes were predicted to be involved in the formation of C19-diterpenoid alkaloids, with 34 DEGs between Z175 and R184. The transcriptome data showed that genes involved in alkaloid biosynthesis are differentially expressed in different varieties, consistent with differences in the accumulation of alkaloids. This expands our understanding of alkaloids, provides a

Table 2
unigenes involved in of salsolinol biosynthesis pathway except glycolysis and shikimate pathway in *A. carmichaelii*.

Enzyme	Enzyme name	Unigene	Unigene length	NR Evalue	Z175 readcount	R184 readcount	Log2 FoldChange	Padj	Significant
CHMU (7)	chorismate mutase	Cluster-219.11902	584	6.60E-84	31.295	32.785	−0.110	0.986	FALSE
		Cluster-219.82304	663	8.60E-72	25.876	17.853	0.487	0.986	FALSE
		Cluster-219.98270	1361	3.00E-111	36.949	19.476	0.893	0.986	FALSE
		Cluster-219.98271	1061	5.70E-78	0.253	2.758	−3.170	0.850	FALSE
		Cluster-219.93045	1224	8.30E-113	572.847	651.924	−0.188	0.976	FALSE
		Cluster-219.67054	1223	2.70E-111	57.315	82.108	−0.522	0.874	FALSE
		Cluster-219.67055	674	8.70E-72	58.800	63.246	−0.108	0.986	FALSE
PAT (5)	bifunctional aspartate aminotransferase and glutamate	Cluster-219.109926	1821	1.40E-188	97.250	137.349	−0.511	0.816	FALSE
		Cluster-219.109927	1786	9.70E-187	103.681	157.760	−0.596	0.456	FALSE
		Cluster-219.139132	542	5.10E-30	4.756	11.616	−1.270	0.696	FALSE
		Cluster-219.132472	698	4.00E-64	3.424	7.438	−1.163	0.881	FALSE
		Cluster-219.132473	922	2.50E-98	8.772	5.868	0.499	0.986	FALSE
TYRC (7)	prephenate dehydrogenase	Cluster-219.51852	2930	5.90E-16	2.893	0.990	1.352	0.986	FALSE
		Cluster-219.57952	2933	5.90E-16	3.743	3.441	0.177	0.990	FALSE
		Cluster-219.53953	3637	7.40E-16	31.361	14.272	1.140	0.975	FALSE
		Cluster-219.53947	3293	1.90E-15	19.069	51.736	−1.414	0.073	FALSE
		Cluster-219.53946	3640	7.40E-16	48.228	38.448	0.337	0.986	FALSE
		Cluster-219.53942	3290	1.90E-15	13.222	24.050	−0.895	0.872	FALSE
		Cluster-219.49397	345	2.80E-10	0.000	0.241	−1.206	0.986	FALSE
POX (23)	polyphenol oxidase	Cluster-219.54921	2241	8.80E-129	1513.885	63.678	4.570	0.140	FALSE
		Cluster-219.54923	1967	6.00E-113	72.495	16.046	2.126	0.347	FALSE
		Cluster-219.7282	570	1.40E-14	38.572	39.573	−0.021	0.996	FALSE
		Cluster-219.55521	2226	1.10E-126	1562.077	246.422	2.663	0.798	FALSE
		Cluster-22,607.0	827	5.30E-71	11.265	0.730	4.873	0.049	TRUE
		Cluster-219.77526	1984	4.40E-196	1700.704	1167.533	0.543	0.810	FALSE
		Cluster-219.14771	2019	2.40E-186	65.567	296.068	−2.177	0.090	FALSE
		Cluster-24,818.0	398	9.20E-13	5.627	0.509	3.136	0.584	FALSE
		Cluster-219.82880	1894	1.80E-106	136.438	21.852	2.647	0.045	TRUE
		Cluster-219.101914	1924	8.00E-187	18.972	70.761	−1.908	0.467	FALSE
		Cluster-219.54919	2117	2.40E-104	434.744	79.363	2.446	0.385	FALSE
		Cluster-219.54918	2044	3.90E-123	1144.021	78.026	3.846	0.000	TRUE
		Cluster-219.54917	2307	1.80E-116	175.363	13.747	3.601	0.003	TRUE
		Cluster-219.21323	607	9.70E-14	10.035	19.630	−0.900	0.898	FALSE
		Cluster-13,152.0	359	4.80E-37	2.112	0.482	1.798	0.986	FALSE
		Cluster-13,152.2	307	8.60E-43	0.760	0.000	1.979	0.986	FALSE
		Cluster-219.2640	668	5.70E-23	4.413	8.486	−0.944	0.983	FALSE
		Cluster-219.46592	716	5.70E-114	47.137	76.689	−0.678	0.957	FALSE
		Cluster-219.50283	885	2.00E-20	476.302	27.677	4.100	0.000	TRUE
		Cluster-17,242.0	1683	4.00E-142	28.044	1.970	4.197	0.013	TRUE
		Cluster-219.70351	626	2.60E-14	276.452	59.749	2.196	0.263	FALSE
		Cluster-219.92608	1191	2.70E-23	404.993	14.361	4.867	0.000	TRUE
		Cluster-219.29862	799	1.40E-15	29.500	18.531	0.661	0.806	FALSE
TYDC (6)	tyrosine decarboxylase	Cluster-219.138055	1681	2.10E-207	0.000	1.018	−2.702	0.986	FALSE
		Cluster-219.138056	789	6.80E-52	17.198	4.987	1.694	0.517	FALSE
		Cluster-219.138245	1746	6.70E-225	89.068	46.722	0.916	0.986	FALSE
		Cluster-219.138244	1722	6.60E-225	15.607	78.879	−2.332	0.275	FALSE
		Cluster-219.138246	1657	2.10E-207	0.000	2.035	−3.668	0.960	FALSE
		Cluster-219.128685	1215	3.60E-108	5.744	1.272	2.083	0.974	FALSE

foundation for future research into the molecular basis of aconitine-type C19-DAs and salsolinol biosynthesis, and sheds light on the synthesis of these acutely toxic components of *Aconitum*.

Funding

This work was supported by the National Key R & D Program (Grant No. 2017YFC1701802 and 2017YFC1700700), the National Chinese Medicine Standardization Project (Grant No. ZYBZH-Y-ZY-45), the Sichuan Provincial Science and Technology Bureau Breeding Program (Grant No. 2016NYZ0036), the Sichuan Science and Technology Bureau Science and Technology support Program (Grant No. 2015SZ0034) and Support of Science and Technology Poverty Alleviation Program of Sichuan Science and Technology Bureau (Grant No. 2018NFP0053).

Declaration of Competing Interest

None.

Acknowledgements

Thanks for the supporting of the six Programs and Jiangyou Fuzi GAP cultivation base.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fitote.2019.104412.

References

[1] L. Ma, R. Gu, L. Tang, Z.-E. Chen, R. Di, C. Long, Important poisonous plants in Tibetan ethnomedicine, Toxins 7 (1) (2015) 138–155.
[2] L.Q. Li, Y. Kadota, Aconitum, in: Z.Y. Wu, P. Raven (Eds.), Flora of China, Science Press and Missouri Botanical Garden Press, Beijing; St. Louis, 2001, pp. 149–222.
[3] X. Wang, X. Li, L. Li, M. Li, Q. Wu, Y. Liu, J. Yang, Y. Jin, Green determination of aconitum alkaloids in *Aconitum carmichaeli* (Fuzi) by an ionic liquid aqueous two-phase system and recovery of the ionic liquid coupled with in situ liquid–liquid microextraction, Anal. Methods-UK 8 (35) (2016) 6566–6572.
[4] P.G. XIAO, F.P. WANG, F. GAO, L.P. YAN, D.L. CHEN, Y. LIU, A pharmaco- phylogenetic study of *Aconitum* L. (Ranunculaceae) from China[J], Acta Phytotaxonomica Sinica 44 (1) (2006) 1–46.
[5] E. Nyirimigabo, Y. Xu, Y. Li, Y. Wang, K. Agyemang, Y. Zhang, A review on phytochemistry, pharmacology and toxicology studies of a conitum, J. Pharm.Pharmacol. 67 (1) (2015) 1–19.
[6] M. Yu, Y.-X. Yang, X.-Y. Shu, J. Huang, D.-B. Hou, *Aconitum carmichaelii* Debeaux, cultivated as a medicinal plant in western China, Genet. Resour. Crop. Evol. 63 (5) (2016) 919–924.
[7] M. Yu, L.-L. Cao, Y.-X. Yang, L.-L. Guan, L.-L. Gou, X.-Y. Shu, J. Huang, D. Liu, H. Zhang, D.-B. Hou, Genetic diversity and marker–trait association analysis for agronomic traits in *Aconitum carmichaelii* Debeaux, Biotechnol. Biotec. Eq. 31 (5) (2017) 905–911.
[8] H. Luo, Z. Huang, X. Tang, J. Yi, S. Chen, A. Yang, J. Yang, Dynamic variation patterns of aconitum alkaloids in daughter root of *Aconitum Carmichaelii* (Fuzi) in the decoction process based on the content changes of nine Aconitum alkaloids by HPLC-MS-MS, Iran, J. Pharm. Res. 15 (1) (2016) 263.
[9] F.P. Wang, X.T. Liang, C20-diterpenoid alkaloids, Alkaloids Chem. Biol. 59 (2002) 1–280.
[10] G. Zhou, L. Tang, X. Zhou, T. Wang, Z. Kou, Z. Wang, A review on phytochemistry and pharmacological activities of the processed lateral root of *Aconitum carmichaelii* Debeaux, J. Ethnopharmacol. 160 (2015) 173–193.
[11] D. Zhao, Y. Shi, X. Zhu, L. Liu, P. Ji, C. Long, Y. Shen, E.J. Kennelly, Identification of potential biomarkers from *Aconitum carmichaelii*, a traditional Chinese medicine, using a metabolomic approach, Planta Med. 84 (06/07) (2018) 434–441.
[12] J. Singhuber, M. Zhu, S. Prinz, B. Kopp, *Aconitum* in traditional Chinese medicine—a valuable drug or an unpredictable risk? J. Ethnopharmacol. 126 (1) (2009) 18–30.
[13] M. Rai, A. Rai, N. Kawano, K. Yoshimatsu, H. Takahashi, H. Suzuki, N. Kawahara, K. Saito, M. Yamazaki, De novo RNA sequencing and expression analysis of *Aconitum carmichaelii* to analyze key genes involved in the biosynthesis of diterpene alkaloids, Molecules 22 (12) (2017) 2155.
[14] F.-P. Wang, Q.-H. Chen, The C19-diterpenoid alkaloids, The Alkaloids: Chem. Biol. 69 (2010) 1–577.
[15] M. Weber, K. Owens, R. Sarpong, Atropurpuran—missing biosynthetic link leading to the hetidine and arcutine C20-diterpenoid alkaloids or an oxidative degradation product? Tetrahedron Lett. 56 (23) (2015) 3600–3603.
[16] E.C. Cherney, P.S. Baran, Terpenoid-alkaloids: their biosynthetic twist of fate and total synthesis, Isr. J. Chem. 51 (3–4) (2011) 391–405.
[17] R.M. Riggin, M.J. McCarthy, P.T. Kissinger, Identification of salsolinol as a major dopamine metabolite in the banana, J. Agr. Food Chem. 24 (1) (1976) 189–191.
[18] W.M. Whaley, T.R. Govindachari, The Pictet-Spengler synthesis of tetrahydroisoquinolines and related compounds, Org. React. 6 (1951) 151–190.
[19] X. Chen, A. Arshad, H. Qing, R. Wang, J. Lu, Y. Deng, Enzymatic condensation of dopamine and acetaldehyde: a salsolinol synthase from rat brain, Biologia 66 (6) (2011) 1183.
[20] D. Chen, L. Hui-yin, W.-L. Song, Study on the constituents of aconite from China (II),chemical composition of rhizoma cyperi, Chin. Herb. Med. 3 (11) (1982) 481–483.
[21] T. KOSUGE, M. YOKOTA, Studies on cardiac principle of aconite root, Chem. Pharm. Bull. 24 (1) (1976) 176–178.
[22] C. Konno, M. Shirasaka, H. Hikino, Cardioactive principle of *Aconitum carmichaeli* Roots1, Planta Med. 35 (2) (1979) 150–155.
[23] W. Jiang, Preliminary clinical observation on the effect of aconite 1 on tardy arrhythmia, Chin. J. Cardiol. 8 (2) (1980) 95–98.
[24] W. Zhang, G. Han, Study on alkaloid composition of aconite from Jiangyou Sichuan province, J. Med. Study 27 (9) (1992) 670.
[25] H.B.J. Grabherr, G. M, M. Yassour, et al., Full-length transcriptome assembly from RNA-Seq data without a reference genome, Nat. Biotechnol. 29 (7) (2011) 644–652.
[26] B. Buchfink, C. Xie, D.H. Huson, Fast and sensitive protein alignment using DIAMOND, Nat. Methods 12 (1) (2015) 59–60.
[27] D. Zhao, Y. Shen, Y. Shi, X. Shi, Q. Qiao, S. Zi, E. Zhao, D. Yu, E.J. Kennelly, Probing the transcriptome of Aconitum carmichaelii reveals the candidate genes associated with the biosynthesis of the toxic aconitine-type C 19-diterpenoid alkaloids, Phytochemistry 152 (2018) 113–124.
[28] P.-J. Zhao, S. Gao, L.-M. Fan, J.-L. Nie, H.-P. He, Y. Zeng, Y.-M. Shen, X.-J. Hao, Approach to the biosynthesis of atisine-type diterpenoid alkaloids, J. Nat. Prod. 72 (4) (2009) 645–649.
[29] M. Kodama, H. Kurihara, S. Itô, A novel rearrangement in C-20 diterpene alkaloids. Formation of bridged bicyclo [4.3. 1] dec-1-ene system, Tetrahedron Lett. 16 (15) (1975) 1301–1304.
[30] K. Chen, P.S. Baran, Total synthesis of eudesmane terpenes by site-selective C–H oxidations, Nature 459 (7248) (2009) 824.
[31] D. Drew, B. Dueholm, C. Weitzel, Y. Zhang, C. Sensen, H. Simonsen, Transcriptome analysis of Thapsia laciniata Rouy provides insights into terpenoid biosynthesis and diversity in Apiaceae, Inter. J Mol. Sci. 14 (5) (2013) 9080–9098.
[32] R.D. Finn, J. Clements, S.R. Eddy, HMMER web server: interactive sequence similarity searching, Nucleic Acids Res 39 (2011) W29–W37 (Nucleic. Acids Research 39(Web Server issue) 29–37).
[33] A. Conesa, S. Götz, J.M. García-Gómez, J. Terol, M. Talón, M. Robles, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, Bioinformatics 21 (18) (2005) 3674–3676.
[34] A. Mortazavi, B.A. Williams, K. Mc Cue, L. Schaeffer, B. Wold, Mapping and quantifying mammalian transcriptomes by RNA-Seq, Nat. Methods 5 (7) (2008) 621–628.
[35] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, BMC Bioinforma. (2011), https://doi.org/10.1186/1471-2105-12-323.
[36] W.H.M.I. Love, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550, https://doi.org/10.1186/s13059-014-0550-8.
[37] W.M.J. Young, D. M, G.K. Smyth, et al., Gene ontology analysis for RNA-seq: accounting for selection bias, Genome Biol. (2010), https://doi.org/10.1186/gb-2010-11-2-r14.
[38] X. Mao, T. Cai, J.G. Olyarchuk, L. Wei, et al., Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary, Bioinformatics 21 (2005) 3787–3793.
[39] B. Canback, S. Andersson, C. Kurland, The global phylogeny of glycolytic enzymes, Pro. Nat. Aca.Sci. 99 (9) (2002) 6097–6102.
[40] K. Kosová, P. Vítámvás, I.T. Prášil, J. Renaut, Plant proteome changes under abiotic stress—contribution of proteomics studies to understanding plant stress response, J. Proteome 74 (8) (2011) 1301–1322.
[41] H. Lee, Y. Guo, M. Ohta, L. Xiong, B. Stevenson, J.K. Zhu, LOS2, a genetic locus required for cold-responsive gene transcription encodes a bifunctional enolase, EMBO J. 21 (11) (2002) 2692–2702.
[42] J.-w. Kim, C.V. Dang, Multifaceted roles of glycolytic enzymes, Trends Biochem. Sci. 30 (3) (2005) 142–150.
[43] M. Rohmer, Mevalonate-independent methylerythritol phosphate pathway for isoprenoid biosynthesis. Elucidation and distribution, Pure Appl. Chem. 75 (2–3) (2003) 375–388.
[44] J. Jin, F. Tian, D.-C. Yang, Y.-Q. Meng, L. Kong, J. Luo, G. Gao, PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants, Nucleic Acids Res 45(D1) (2017) D1040–D1045 gkw982.
[45] A. Kumar, S. Kumar, S. Bains, V. Vaidya, B. Singh, R. Kaur, J. Kaur, K. Singh, De novo transcriptome analysis revealed genes involved in flavonoid and vitamin C biosynthesis in *Phyllanthus emblica* (L.), Front. Plant Sci. 7 (2016) 1610.
[46] L. Yang, G. Ding, H. Lin, H. Cheng, Y. Kong, Y. Wei, X. Fang, R. Liu, L. Wang, X. Chen, Transcriptome analysis of medicinal plant Salvia miltiorrhiza and identification of genes related to tanshinone biosynthesis, PLoS One 8 (11) (2013) e80464.
[47] T. Pal, N. Malhotra, S.K. Chanumolu, R.S. Chauhan, Next-generation sequencing (NGS) transcriptomes reveal association of multiple genes and pathways contributing to secondary metabolites accumulation in tuberous roots of Aconitum

heterophyllum wall, Planta 242 (1) (2015) 239–258.
[48] K. Jørgensen, A.V. Rasmussen, M. Morant, A.H. Nielsen, N. Bjarnholt, M. Zagrobelny, S. Bak, B.L. Møller, Metabolon formation and metabolic channeling in the biosynthesis of plant natural products, Curr. Opin. Plant Biol. 8 (3) (2005) 280–291.
[49] B.S. Winkel, Metabolic channeling in plants, Annu. Rev. Plant Biol. 55 (2004) 85–107.
[50] D. Tholl, S. Lee, Terpene specialized metabolism in *Arabidopsis thaliana*, The Arabidopsis Book/American Society of Plant Biologists 9, (2011).
[51] D. Tholl, F. Chen, J. Petri, J. Gershenzon, E. Pichersky, Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from *Arabidopsis* flowers, Plant J. 42 (5) (2005) 757–771.
[52] N. Dudareva, F. Negre, D.A. Nagegowda, I. Orlova, Plant volatiles: recent advances and future perspectives, Crit. Rev. Plant Sci. 25 (5) (2006) 417–440.
[53] R.C. Misra, A. Garg, S. Roy, C.S. Chanotiya, P.G. Vasudev, S. Ghosh, Involvement of an ent-copalyl diphosphate synthase in tissue-specific accumulation of specialized diterpenes in Andrographis paniculata, Plant Sci. 240 (2015) 50–64.
[54] A. Garg, L. Agrawal, R.C. Misra, S. Sharma, S. Ghosh, *Andrographis paniculata* transcriptome provides molecular insights into tissue-specific accumulation of medicinal diterpenes, BMC Genomics 16 (1) (2015) 659.
[55] V. Tzin, G. Galili, The Biosynthetic Pathways for Shikimate and Aromatic Amino Acids in *Arabidopsis thaliana*, the Arabidopsis Book/American Society of Plant Biologists 8, (2010).
[56] M. Bischoff, A. Schaller, F. Bieri, F. Kessler, N. Amrhein, J. Schmid, Molecular characterization of tomato 3-dehydroquinate dehydratase-shikimate: NADP oxidoreductase, Plant Physiol. 125 (4) (2001) 1891–1900.
[57] S.A. Singh, D. Christendat, Structure of *Arabidopsis* dehydroquinate dehydratase-shikimate dehydrogenase and implications for metabolic channeling in the shikimate pathway, Biochem. 45 (25) (2006) 7787–7796.
[58] K. Kasai, T. Kanno, M. Akita, Y. Ikejiri-Kanno, K. Wakasa, Y. Tozawa, Identification of three shikimate kinase genes in rice: characterization of their differential expression during panicle development and of the enzymatic activities of the encoded proteins, Planta 222 (3) (2005) 438–447.
[59] H. Maeda, N. Dudareva, The shikimate pathway and aromatic amino acid biosynthesis in plants, Ann. Rev. Plant Biol. 63 (1) (2012) 73.
[60] W. Wen, M. Jin, K. Li, H. Liu, Y. Xiao, M. Zhao, S. Alseekh, W. Li, A.E.L.F. De, Y. Brotman, An integrated multi-layered analysis of the metabolic networks of different tissues uncovers key genetic components of primary metabolism in maize, plant J, Cell. Mol. Biol. (2018), https://doi.org/10.1111/tpj.13835.
[61] Q. Han, R.S. Phillips, J. Li, Aromatic amino acid metabolism, front, Mol. Biosci. 6 (22) (2019).
[62] M.L. Sullivan, Beyond brown: polyphenol oxidases as enzymes of plant specialized metabolism, Front. Plant Sci. 5 (2015) 783.
[63] Z.J. Gao, X.H. Han, X.G. Xiao, Purification and characterisation of polyphenol oxidase from red Swiss chard (Beta vulgaris subspecies cicla) leaves, Food Chem. 117 (2) (2009) 342–348.
[64] Y. Zhou, X. Wang, W. Wang, H. Duan, De novo transcriptome sequencing-based discovery and expression analyses of verbascoside biosynthesis-associated genes in Rehmannia glutinosa tuberous roots, Mol. Breeding 36 (10) (2016) 139.
[65] M.M. Sojo, E. Nuñez-Delicado, A. Sánchez-Ferrer, F. García-Carmona, Oxidation of salsolinol by banana pulp polyphenol oxidase and its kinetic synergism with dopamine, J. Agr. Food Chem. 48 (11) (2000) 5543–5547.
[66] X. Chen, X. Zheng, S. Ali, M. Guo, R. Zhong, Z. Chen, Y. Zhang, H. Qing, Y. Deng, Isolation and sequencing of salsolinol synthase, an enzyme catalyzing salsolinol biosynthesis, ACS Chem. Neurosci. 9 (6) (2018) 1388–1398.