

## 藏药解吉基原植物转录组及其环烯醚萜苷类生物合成相关基因的探究

倪梁红<sup>1</sup>, 赵志礼<sup>1\*</sup>, 吴靳荣<sup>1</sup>, 嘎 务<sup>2</sup>, 米 玛<sup>2</sup>

(1. 上海中医药大学中药学院, 上海 201203; 2. 西藏藏医药大学, 西藏 拉萨 850000)

**摘要:** 藏药解吉 (ཇེ་ཇི), 来源为龙胆科龙胆属 *Gentiana* 秦艽组 (Sect. *Cruciata*) 多种高山植物; 可分为解吉那保 (具蓝紫色花的) 与解吉嘎保 (具白色花的) 两个品种。首次以解吉那保基原植物之一——长梗秦艽 *Gentiana waltonii* Burk. 与解吉嘎保基原植物之一——粗壮秦艽 *Gentiana robusta* King ex Hook. f. 为模式植物, 分别构建其根、茎、叶、花转录组数据库。对原始数据进行除杂, 将质控后得到的所有高质量序列进行 *de novo* 拼接, 分别得到长梗秦艽 79 455 条、粗壮秦艽 78 466 条 unigene, 平均长度分别为 834 bp 和 862 bp。根据 GO 功能分类, 可将 unigene 分为 3 大类 65 分支; 根据 KOG 功能可将 unigene 分为 25 类; KEGG 代谢通路分析分别发现长梗秦艽和粗壮秦艽中 315 条、340 条 unigene 参与 20 个次生代谢标准通路中。环烯醚萜苷类为藏药解吉及中药秦艽类、龙胆类、当药及青叶胆等药材的指标性成分及活性成分之一, 分析结果显示: 分别有 80 条、57 条 unigene 参与编码环烯醚萜苷类合成途径中的 24 个关键酶, 且在不同部位中的表达存在差异; 还发现多个 SSR 和 SNP、InDel 位点。本工作可为藏药材基原植物遗传图谱构建, 分子标记开发, 功能基因掌握, 有效成分积累规律、道地性形成机制、藏药材品种科学内涵的探讨, 分子育种及濒危藏药高山植物就地保护等提供基础科学资料。

**关键词:** 长梗秦艽; 粗壮秦艽; 转录组; 环烯醚萜苷类; 生物合成; 代谢通路

中图分类号: R931 文献标识码: A 文章编号: 0513-4870(2019)05-0944-10

## Analysis of transcriptomes to explore genes contributing to iridoid biosynthesis in *Gentiana waltonii* and *Gentiana robusta* (Gentianaceae)

NI Liang-hong<sup>1</sup>, ZHAO Zhi-li<sup>1\*</sup>, WU Jin-rong<sup>1</sup>, GAAWE Dorje<sup>2</sup>, MI Ma<sup>2</sup>

(1. School of Pharmacy, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China;

2. Tibetan Traditional Medical College, Lhasa 850000, China)

**Abstract:** The roots and flowers of *Gentiana waltonii* and *Gentiana robusta* are used as Tibetan herb Jie-Ji in traditional Tibetan medicine, with iridoids as the main active ingredient and index components. To study the pathway of iridoid biosynthesis, roots, stems, leaves and flowers of *G. waltonii* and *G. robusta* were subjected to a high-throughput transcriptomic sequencing analysis by Illumina HiseqXTen. After removing insignificant reads and *de novo* splicing, 79 455 and 78 466 unigenes were obtained from *G. waltonii* and *G. robusta* respectively, with average length as 834 bp and 862 bp. The unigene GO functions could be divided into three categories of 65 branches. The unigenes were aligned in KOG database and were classified into 25 classes according to function. In KEGG database, 315 and 340 unigenes of *G. waltonii* and *G. robusta* were implicated in 20 standard secondary metabolic pathways, respectively. Furthermore, 80 and 57 unigenes of the two species were analyzed to encode 24 key enzymes in the pathway related to iridoid biosynthesis. There were differences in gene expression among

收稿日期: 2019-01-28; 修回日期: 2019-02-23.

基金项目: 国家自然科学基金青年科学基金资助项目 (81503354); 国家自然科学基金资助项目 (81173654).

\*通讯作者 Tel: 86-21-51322202, E-mail: zhilzhao@sohu.com

DOI: 10.16438/j.0513-4870.2019-0082

different organs. Based on sequence data, significant amounts of SSRs, SNPs and InDels were detected in each dataset. This study provides a platform for further development of molecular markers, excavation of functional genes, and research into metabolic pathways and their regulatory mechanism within *G. waltonii* and *G. robusta*.

**Key words:** *Gentiana waltonii*; *Gentiana robusta*; transcriptome; iridoids; biosynthesis; metabolism pathway

解吉(ཉེ་མེ), 传统藏药之一, 常用于治疗“赤巴病”<sup>[1]</sup>。根据基原植物花色不同, 历史上又分为解吉那保(ཉེ་མེ་ནག་པོ།; 具蓝紫色花的)与解吉嘎保(ཉེ་མེ་དཀར་པོ།; 具白色花的)两个品种<sup>[2]</sup>。据文献资料及多年民族植物学考察, 其来源为龙胆科龙胆属 *Gentiana* 秦艽组 (Sect. *Cruciata*) 多种高山植物, 花、根入药, 散风祛湿, 清热利胆<sup>[3,4]</sup>。

课题组前期开展了秦艽组药用植物的分类学鉴定及遗传背景分析等基础性工作<sup>[4-7]</sup>, 确定了藏药解吉基原植物之一的粗壮秦艽 *Gentiana robusta* King ex Hook. f. 为一自然杂交物种; 基于相关分析, 推断其母系来源可能为另一高山药用植物麻花秦艽 *Gentiana straminea* Maxim.<sup>[8]</sup>。还注意到一些物种 ITS 区及叶绿体片段在基因组内存在丰富的多样性(多种基因型), 构建物种 DNA 条形码需要更为丰富的遗传背景资料及个性化的物种鉴定策略<sup>[9]</sup>。同时, 复杂的遗传背景及其所含化学成分的丰富多样, 为寻找安全、高效的临床药物提供了可能性。

藏医药学具有悠久的发展历史及丰富的科学与藏族文化内涵, 根据药物外观颜色或基原植物花色不同可再分为不同品种是其鲜明特色之一。如藏药“巴莪”有“巴莪色保”(དཔལ་ལོ་ལོ་ལོ།; “黄色的”)及“巴莪嘎保”(དཔལ་ལོ་དཀར་པོ།; “白色的”)之分<sup>[2,10]</sup>; 藏药“庞阿”有“庞阿那保”(ལོང་ལ་ནག་པོ།; “蓝紫色的”)、“庞阿嘎保”(ལོང་ལ་དཀར་པོ།; “白色的”)及“庞阿玛保”(ལོང་ལ་དམར་པོ།; “红色的”)3个品种之分<sup>[3]</sup>等。开展相关基础性研究工作, 分析不同品种间疗效的异同, 深入探讨其科学内涵, 科学意义不言而喻。

环烯醚萜苷类(iridoids)广泛分布于龙胆科(Gentianaceae)、玄参科(Scrophulariaceae)及茜草科(Rubiaceae)等双子叶植物中, 具有抗病毒、抗炎、抗氧化、抗肿瘤、抗菌及保肝利胆等多种药理作用<sup>[11]</sup>, 为中药秦艽类、龙胆类、当药及青叶胆等药材的指标性成分及活性成分之一。如2015版《中华人民共和国药典》秦艽项下“龙胆苦苷和马钱苷酸”, 龙胆项下“龙胆苦苷”, 当药项下“当药苷和獐牙菜苦苷”及青叶胆项下“獐牙菜苦苷”的含量测定等<sup>[12]</sup>。

植物转录组是特定组织或细胞在某一功能状态下

转录出来的所有RNA的总和, 狭义上的转录组一般指所有mRNA的集合<sup>[13]</sup>。高通量转录组测序技术现已广泛应用于药用植物转录组分析, 可全面快速地获取研究对象在某一状态下基因转录信息, 从中挖掘重要功能基因, 揭示不同生物学性状的分子机制, 为研究中药生长发育及次生代谢的分子机制提供重要的基因资源<sup>[14]</sup>。

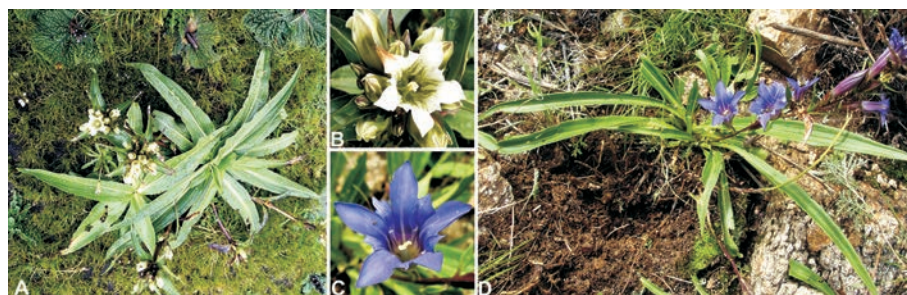
为此, 在前期藏药解吉品种整理等工作基础上, 分别以解吉那保基原植物之一——长梗秦艽 *Gentiana waltonii* Burk. 与解吉嘎保基原植物之一——粗壮秦艽 *G. robusta* King ex Hook. f. 为模式植物, 应用 Illumina HiSeqXTen 测序技术构建其转录组数据库, 揭示各自转录组的整体表达特征, 比较分析异同, 寻找分子标记, 并挖掘环烯醚萜苷类生物合成通路关键酶基因; 期为藏药材基原植物遗传图谱构建, 分子标记开发, 功能基因掌握, 有效成分累积规律、道地性形成机制、藏药材品种科学内涵的探讨, 分子育种及濒危藏药高山植物就地保护等提供基础科学资料。

## 材料与方法

**民族植物学考察及分类学鉴定** 与西藏藏医药大学藏族同行合作开展藏药“解吉”的考察工作。凭证标本及实验样品采集信息: 粗壮秦艽, 西藏墨竹工卡县, 采集号: 赵志礼, 倪梁红 Z2018XZ002, 地理坐标 N 29°41.315', E 92°06.798', 海拔 4 315 m。长梗秦艽, 西藏墨竹工卡县, 采集号: 赵志礼, 倪梁红 Z2018XZ001, 地理坐标 N 29°43.808', E 91°59.067', 海拔 4 050 m (图 1)。取正常生长无病害植株, 连土采挖后鲜活状态下运至实验室, 纯水清洗, 无菌吸水纸吸干表面水分, 迅速将根、茎、叶、花各部位分别切成小块, 液氮速冻, 备用。

植物经上海中医药大学赵志礼教授鉴定为龙胆科龙胆属 *Gentiana* 秦艽组 (Sect. *Cruciata*) 粗壮秦艽 *Gentiana robusta* King ex Hook. f. 及长梗秦艽 *Gentiana waltonii* Burk.; 凭证标本存放于上海中医药大学中药学院药用植物标本室。

**RNA 的提取分离、文库构建和转录组测序** 使用 Total RNA Extractor (Trizol) 试剂盒(上海生工)分别提取长梗秦艽和粗壮秦艽样品根、茎、叶和花的总



**Figure 1** A: Habit of *G. robusta*; B: White flowers of *G. robusta*; C: A dark blue flower of *G. waltonii*; D: Habit of *G. waltonii*

RNA, 检测 RNA 浓度及完整性。使用磁力架 (上海生物工程股份有限公司) 以磁珠法分离 mRNA, 将 mRNA 片段化, 完成双链 cDNA 合成、末端修复、末端 dA-Tailing、接头连接、连接产物纯化和片段大小分选、文库扩增 (15 个循环), 回收 DNA 后精确定量, 按数据比例混合上机。在 Illumina HiSeqXTen 测序平台进行  $2 \times 150$  bp 测序。

**测序数据处理及生物信息学分析** 原始图像数据文件经碱基识别 (base calling) 分析转化为原始测序序列 (sequenced reads), 以 fastq 文件格式存储。对原始数据质量值等信息进行统计, 并使用 FastQC 对样本的测序数据质量进行可视化评估, 过滤除去含有带接头的、低质量的序列, 得到 clean 数据。使用 Trinity 软件将 clean 数据 *de novo* 组装成转录本 (transcript)<sup>[15]</sup>, 对转录本去冗余, 取每个转录本聚类中最长的转录本作为非重复序列基因 (unigene)。

对转录数据使用 BlastX 软件 ( $E$  值  $< 1 \times 10^{-5}$ ) 将 unigene 序列与 CDD、NR、NT、Swissprot、KOG (cluster of orthologous groups of proteins) 等数据库比对, 进行功能注释和分类; 再对 unigene 序列进行 GO (gene ontology) 功能注释和分类, 并与 KEGG (kyoto encyclopedia of genes and genomes) 数据进行比对, 分析相关的代谢通路。

使用 MISA<sup>[16]</sup> 对 unigene 序列进行 SSR (simple sequence repeat) 检测, 参数设置如下: 单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸和六核苷酸 SSR 的重复次数分别至少为 10、6、5、5、5、5, 间隔不超过 100 bp 的复合型 SSR。使用 BCFtools<sup>[17]</sup> 进行 SNP/InDel calling, 抽提各样本中的 SNP 和 InDel, 并对其根

据以下条件进行过滤: ① 质量值大于 20; ② 覆盖度大于 8。

**环烯醚萜苷类生物合成基因筛选** 根据文献中环烯醚萜苷类化合物的生物合成途径<sup>[18-20]</sup>, 结合 NR 和 KEGG 数据库注释结果和已有基因信息, 挖掘本转录组中与环烯醚萜苷类合成相关的 unigene, 以 TPM (transcripts per million) 值, 进行表达量的统计, TPM 值同时考虑了测序深度和基因长度以及样本对 reads 计数的影响, 是常用的基因表达水平估算方法<sup>[21]</sup>。

## 结果与分析

### 1 转录组数据组装

对长梗秦艽和粗壮秦艽转录组测序及 raw reads 片段除杂后, 长梗秦艽根、茎、叶和花 4 个部位分别获得 46 701 370、42 514 714、44 696 544 及 45 408 678 个高质量 reads 片段; 粗壮秦艽根、茎、叶和花 4 个部位分别获得 46 749 340、43 927 220、48 944 798 及 43 556 026 个高质量 reads 片段。序列上传 SRA 数据库, 登录号: 长梗秦艽 (SRR8185390-SRR8185393), 粗壮秦艽 (SRR8381079-SRR8381082)。将质控后得到的所有高质量序列进行 *de novo* 拼接。共拼接得到长梗秦艽 unigene 79 455 条, 平均长度为 834 bp; 拼接得到粗壮秦艽 unigene 78 466 条, 平均长度为 862 bp (表 1)。

### 2 转录组 unigenes 功能注释

将 unigene 与 NR、NT、GO、KEGG、Swissprot 及 KOG 等数据库通过 BLAST 进行比对与注释, 长梗秦艽与粗壮秦艽分别有 51.97% 和 52% 的 unigene 在至少一种数据库中得到注释 (表 2)。

长梗秦艽与粗壮秦艽的 unigene 在 NR 数据库比

**Table 1** Summary of the transcript statistics generated from *G. waltonii* and *G. robusta*

Species		Number	$\geq 500$ bp	$\geq 1\ 000$ bp	N50	N90	Max length/bp	Min length/bp	Total length/bp	Average length/bp
<i>G. waltonii</i>	Transcript	209 248	125 697	74 439	1 579	428	12 762	201	208 537 252	997
	Unigene	79 455	26 490	24 425	1 499	309	12 762	201	66 272 371	834
<i>G. robusta</i>	Transcript	219 633	135 726	82 686	1 672	454	15 691	201	230 402 981	1 049
	Unigene	78 466	37 123	21 476	1 559	321	15 691	201	67 651 358	862

**Table 2** Statistics of unigene functional annotation

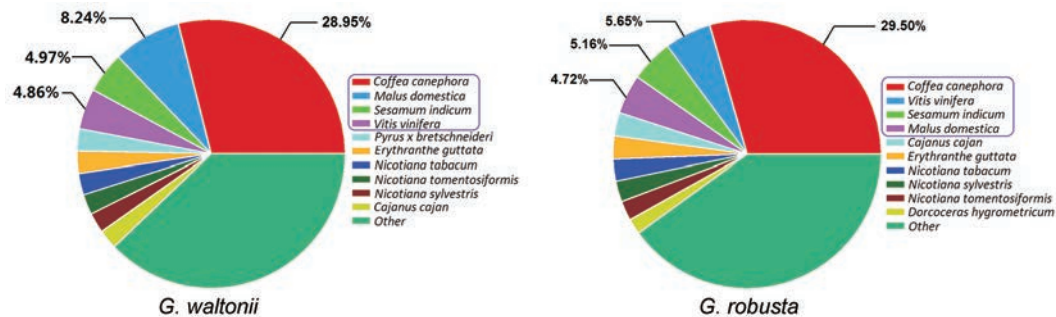
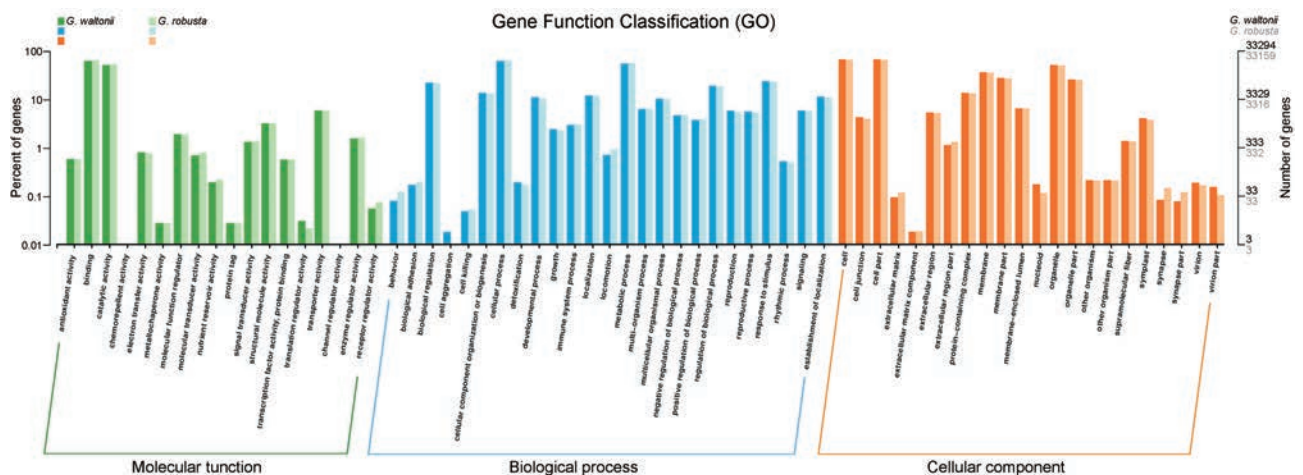
Database (annotated)	<i>G. waltonii</i>		<i>G. robusta</i>	
	No. of unigenes	Percentage %	No. of unigenes	Percentage %
CDD	23 523	29.61	23 514	29.97
KOG	15 790	19.87	15 395	19.62
NR	39 474	49.68	39 100	49.83
NT	19 832	24.96	17 512	22.32
PFAM	17 327	21.81	17 253	21.99
Swissprot	28 670	36.08	28 455	36.26
TrEMBL	39 001	49.09	38 753	49.39
GO	33 294	41.9	33 159	42.26
KEGG	3 637	4.58	3 521	4.49
At least one database	41 289	51.97	40 802	52
All database	2 127	2.68	1 985	2.53
Total genes	79 455	100	78 466	100

时,相似序列匹配度最高的前10个物种中,有9个一致,其中占比最高的前4个物种分别为咖啡 *Coffea canephora* (28.95% 和 29.50%)、苹果 *Malus domestica* (8.24% 和 4.72%)、芝麻 *Sesamun indicum* (4.97% 和 5.16%) 和葡萄 *Vitis vinifera* (4.86% 和 5.65%), 其余物种均小于3% (图2)

长梗秦艽和粗壮秦艽转录组中, 分别有33 294条及33 159条 unigene 在 GO 数据库中成功注释, 根据 GO 功能分为生物过程 (biological process)、细胞组分 (cellular component) 和分子功能 (molecular function)

3个大类共65个分支(图3)。对各类基因数量统计结果显示:长梗秦艽、粗壮秦艽聚集的大类相一致。在生物过程大类中,主要聚集于细胞过程 (cellular process, 分别为20 272条、20 326条)和代谢过程 (metabolic process, 分别为17 920条、17 972条);在细胞组分大类中,主要聚集于细胞 (cell, 分别为21 827条、21 275条)和细胞组分 (cell part, 分别为21 790条、21 234条);在分子功能大类中,主要聚集于结合蛋白 (binding, 分别为20 048条、20 523条)和催化活性 (catalytic activity, 分别为16 646条、16 782条)。

应用 KOG 数据库进行功能预测和分类统计,长梗秦艽、粗壮秦艽分别共有15 790条、15 395条 unigene 得到注释。根据功能可分为25类,涉及大部分生命活动(图4)。其中,翻译后修饰 (posttranslational modification)、蛋白质周转 (protein turnover)、分子伴侣 (chaperones) 是最大类别,长梗秦艽、粗壮秦艽分别包含有2 114条、2 107条 unigene;其次是信号传导机制 (signal transduction mechanisms, 分别为2 010条、2 046条)和一般功能预测 (general function prediction only, 分别为1 989条、1 871条);最少的是细胞外结构 (extracellular structures, 分别为52条、58条)和细胞运动性 (cell motility, 分别为4条、5条)。

**Figure 2** Species distribution of transcriptomic unigenes against NR database**Figure 3** GO classification of transcriptomic unigenes

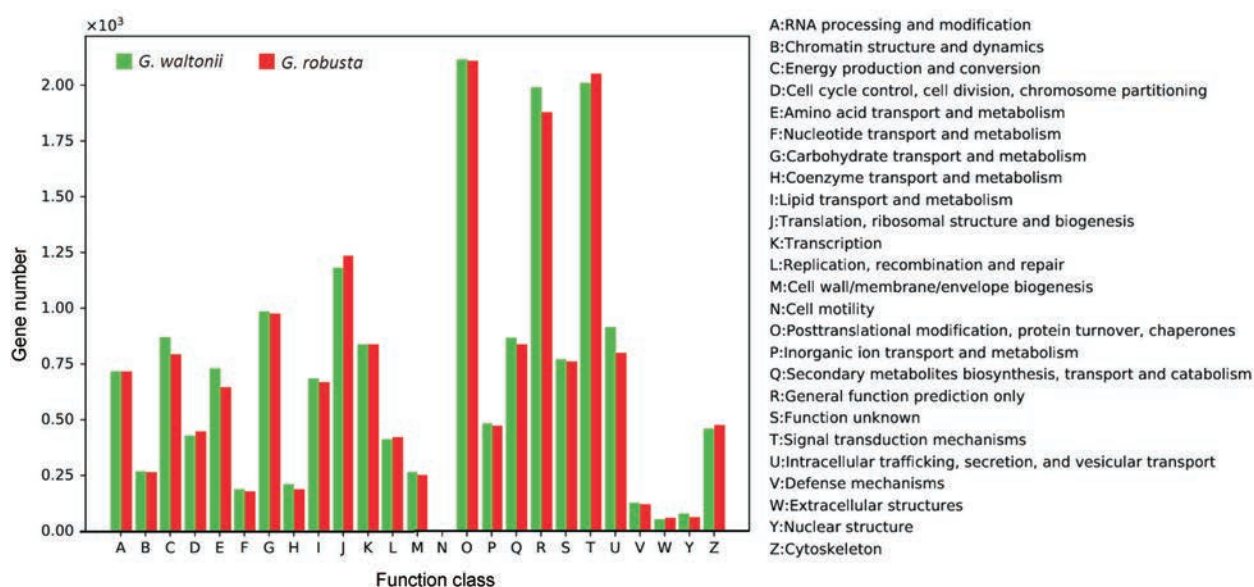


Figure 4 KOG annotation distribution of transcriptomic unigenes

长梗秦艽和粗壮秦艽 unigene 参与的 KEGG 代谢通路可分为 4 个分支: 细胞过程 (cellular processes)、环境信息处理 (environmental information processing)、遗传信息处理 (genetic information processing) 及代谢 (metabolism) (图 5), 包括 23 个中类、212 条代谢通路。长梗秦艽、粗壮秦艽获得注释量前 5 的通路分别是: 核糖体途径, 分别有 307 条 (4.49%)、261 条 (3.91%); 碳代谢, 分别有 175 条 (2.56%)、174 条 (2.61%); 剪接体, 分别有 161 条 (2.35%)、148 条 (2.22%); 内质网蛋白加工, 分别有 159 条 (2.32%)、157 条 (2.35%); 氨基酸生物合成, 分别有 156 条 (2.28%)、152 条 (2.28%)。其余主要

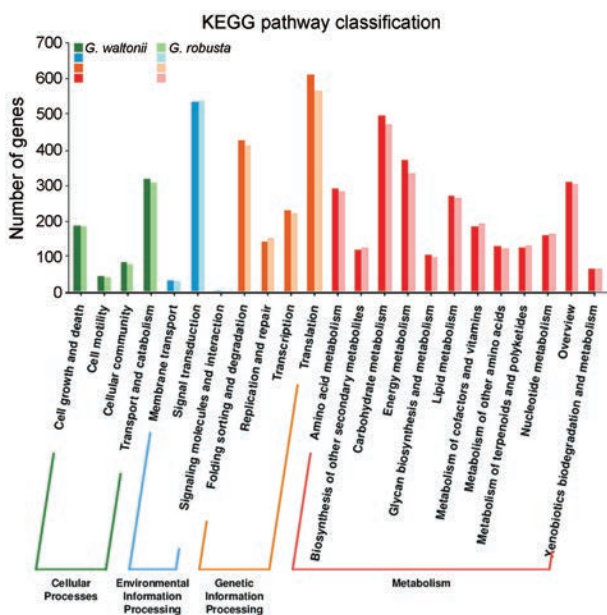


Figure 5 KEGG classification of assembled unigenes

富集于淀粉蔗糖代谢、植物激素信号、转导 RNA 转运、氧化磷酸化等代谢途径。

高等植物含有多种次生代谢产物, 与其药用等功效密切相关。在长梗秦艽和粗壮秦艽转录组中分别发现 315 条和 340 条 unigene 参与到 20 个 KEGG 次生代谢标准通路中。其中, 苯丙素的生物合成 (phenylpropanoid biosynthesis) 最多, 分别有 68 条、73 条; 其次为萜类化合物骨架合成 (terpenoid backbone biosynthesis), 分别有 36 条、38 条。各种萜类化合物的合成 (monoterpenoid, diterpenoid, sesquiterpenoid and triterpenoid biosynthesis) 分别有 26 条、34 条; 黄酮类成分的合成 (flavonoid, isoflavonoid, flavone and flavonol biosynthesis) 分别有 22 条、25 条; 此外, 分别还有 28 条、27 条 unigene 注释到多种生物碱类合成 (indoleisoquinoline, tropane, piperidine and pyridine alkaloid biosynthesis) (表 3)。

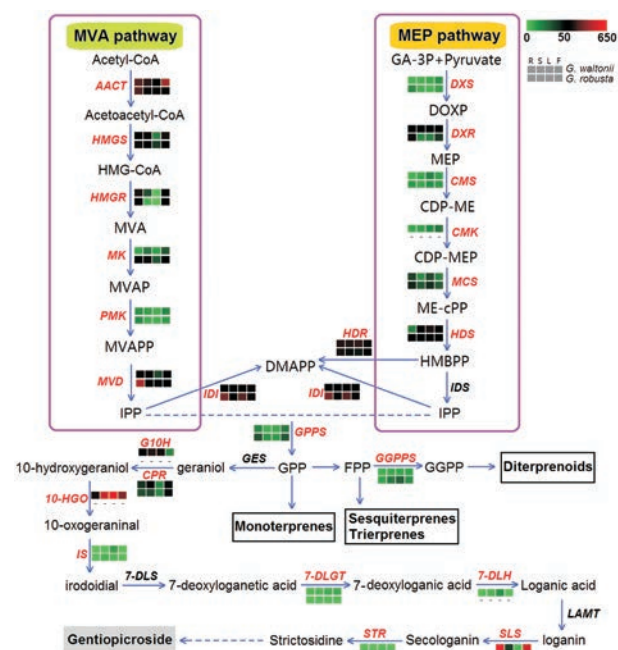
### 3 环烯醚萜苷类合成相关基因

环烯醚萜苷类化合物的生物合成途径可通过中间体的生成、环烯醚萜类骨架构成及修饰等步骤完成<sup>[18-20]</sup> (图 6)。分析结果显示: 长梗秦艽和粗壮秦艽转录组中分别有 80 条、57 条 unigene 参与编码其中 24 个关键酶 (表 4)。

公认环烯醚萜苷中间体生成途径有两条, 一条为甲羟戊酸 (mevalonic acid, MVA) 途径, 该途径在胞质中进行, 关键酶有乙酰-CoA 乙酰转移酶 (AACT)、HMG-CoA 合酶 (HMGS)、HMG-CoA 还原酶 (HMGR)、MVA 激酶 (MK)、MVP 激酶 (PMK) 及 MVPP 脱羧酶 (MVD); 另一条为 2-C-甲基-D-赤藓醇 4-磷酸 (2-meth-

**Table 3** Secondary metabolism KEGG pathway analysis of transcriptomic unigenes

No.	KEGG pathway	Pathway ID	No. of unigenes	
			<i>G. waltonii</i>	<i>G. robusta</i>
1	Phenylpropanoid biosynthesis	ko00940	68	73
2	Terpenoid backbone biosynthesis	ko00900	36	38
3	Ubiquinone and other terpenoid-quinone biosynthesis	ko00130	30	38
4	Steroid biosynthesis	ko00100	30	29
5	Carotenoid biosynthesis	ko00906	29	24
6	Flavonoid biosynthesis	ko00941	18	19
7	Nicotinate and nicotinamide metabolism	ko00760	16	17
8	Tropane, piperidine and pyridine alkaloid biosynthesis	ko00960	14	13
9	Diterpenoid biosynthesis	ko00904	13	16
10	Isoquinoline alkaloid biosynthesis	ko00950	13	13
11	Zeatin biosynthesis	ko00908	12	13
12	Sesquiterpenoid and triterpenoid biosynthesis	ko00909	10	14
13	Stilbenoid, diarylheptanoid and gingerol biosynthesis	ko00945	6	9
14	Brassinosteroid biosynthesis	ko00905	7	8
15	Flavone and flavonol biosynthesis	ko00944	4	5
16	Monoterpenoid biosynthesis	ko00902	3	4
17	Limonene and pinene degradation	ko00903	3	3
18	Caffeine metabolism	ko00232	2	2
19	Indole alkaloid biosynthesis	ko00901	1	1
20	Isoflavonoid biosynthesis	ko00943	0	1



**Figure 6** Putative pathways for gentiopicroside biosynthesis. Genes found in this study are shown between the reactions. The expression of unigenes in the root (R), stem (S), leaf (L) and flower (F) is shown by heatmap, mapped using TPM values

yl-D-erythritol-4-phosphate, MEP) 途径, 在质体中进行, 关键酶包括 1-脱氧-D-葡萄糖-5-磷酸合成酶 (DXS)、1-脱氧-D-葡萄糖-5-磷酸还原异构酶 (DXR)、2-甲基-D-赤藓糖醇-4-磷酸胞苷酰基转移酶 (CMS)、4-二磷酸胞苷-2-甲基-D-赤藓糖醇激酶 (CMK)、2-C-甲基-赤藓糖醇-2,4-环焦磷酸合成酶 (MCS)、4-羟基-3-甲基丁烯-2-烯基-1-二磷酸合酶 (HDS) 及 4-羟基-3-甲基丁烯-2-烯

基-1-二磷酸还原酶 (HDR); 两条途径的产物异戊烯焦磷酸 (isopentenyl pyrophosphate, IPP) 和二甲基丙烯基焦磷酸 (dimethylallyl pyrophosphate, DMAPP) 之间的转换由 IPP 异构酶 (IDI) 来完成。在长梗秦艽和粗壮秦艽转录组中, 除了粗壮秦艽的 CMK 基因未被注释到, 其他以上相关基因均有注释。基于 TPM 值统计, 这些基因表达量存在较明显的差异, 其中 MVA 途径中的 AACT、MVD, MEP 途径中的 HDR 及 IDI 基因表达量相对较高。

IPP 和 DMAPP 在香叶基焦磷酸合酶 (GPPS) 催化下缩合成香叶基焦磷酸 (geranyl diphosphate, GPP), GPP 通过不同的代谢方向流向单萜、二萜、三萜等合成途径。对于环烯醚萜苷的合成, 后续已知的酶基因主要有香叶醇合成酶 (geraniol synthase, GES)、香叶醇-10-羟化酶 (G10H)、细胞色素 P450 还原酶 (CPR)、10-羟基香叶醇氧化还原酶 (10-HGO)、环烯醚萜合酶 (IS)、7-去氧番木鳖酸合成酶 (7-deoxyloganic acid synthase, 7-DLS)、7-去氧番木鳖酸葡萄糖转移酶 (7-DLGT)、7-去氧番木鳖酸羟化酶 (7-DLH)、番木鳖酸甲基转移酶 (loganic acid methyltransferase, LAMT)、裂环马钱子苷合成酶 (SLS) 和异胡豆苷合成酶 (STR)。在长梗秦艽和粗壮秦艽转录组中, GES、7-DLS 和 LAMT 3 个基因没有被注释到, 以及长梗秦艽中的 10-HGO, 粗壮秦艽中的 G10H 和 7-DLH 未有注释, 其他相关基因均有不同数量的 unigene 成功注释。基于 TPM 值统计, 在这两个物种中表达量最高的均为根和花中 SLS 基因, 达 345.94~633.93, 而该基因在茎和叶中的表达较低。

**Table 4** Unigenes involved in iridoid biosynthesis

Pathway	Gene name	Abbr.	EC	<i>G. waltonii</i>						<i>G. robusta</i>					
				No. of unigenes			Average TPM			No. of unigenes			Average TPM		
				Root	Stem	Leaf	Root	Stem	Leaf	Root	Stem	Leaf	Root	Stem	Leaf
MVA	Acetyl-CoA C-acetyltransferase	AACT	2.3.1.9	1	242.08	165.56	73.50	387.57	1	229.36	105.75	95.90	93.36		
	Hydroxymethylglutaryl-CoA synthase	HMGS	2.3.3.10	7	73.36	108.12	19.35	142.94	2	119.45	60.21	37.47	83.56		
	Hydroxymethylglutaryl-CoA reductase (NADPH)	HMGR	1.1.1.34	1	142.65	32.44	0.58	82.27	1	88.02	5.66	0.24	53.66		
	Mevalonate kinase	MK	2.7.1.36	2	9.25	22.39	10.90	30.21	1	72.77	111.96	31.15	65.35		
	Phosphomevalonate kinase	PMK	2.7.4.2	5	15.05	11.37	6.81	17.13	3	16.41	7.09	6.01	11.83		
	Diphosphomevalonate decarboxylase	MVD	4.1.1.33	3	95.71	68.34	34.48	73.95	1	403.19	76.08	47.36	60.89		
	1-Deoxy-D-xylulose-5-phosphate synthase	DXS	2.2.1.7	4	5.37	5.79	7.62	14.48	4	16.64	6.62	4.60	11.10		
	1-Deoxy-D-xylulose-5-phosphate reductoisomerase	DXR	1.1.1.267	1	85.83	75.52	65.98	72.63	1	118.66	18.11	20.45	38.25		
	2-C-methyl-D-erythritol 4-phosphate cytidyltransferase	CMS	2.7.7.60	2	5.08	8.35	18.55	10.82	1	11.85	8.24	14.65	12.49		
	4-Diphosphocytidylyl-2-C-methyl-D-erythritol kinase	CMK	2.7.1.148	1	8.72	10.85	16.10	22.31	0	-	-	-	-		
MEP	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	MCS	4.6.1.12	1	35.37	16.44	33.91	19.87	2	41.87	29.63	40.60	29.11		
	4-Hydroxy-3-methylbut-2-enyl-diphosphate synthase	HDS	1.17.7.3	1	20.09	84.57	140.98	91.42	1	88.65	50.91	81.88	80.25		
	4-Hydroxy-3-methylbut-2-en-1-yl diphosphate reductase	HDR	1.17.7.4	1	155.58	155.77	178.16	118.87	1	137.53	46.51	44.13	90.06		
	Isopentenyl-diphosphate delta-isomerase	IDI	5.3.3.2	3	65.83	75.85	90.61	90.39	1	303.91	94.13	242.89	110.76		
	Geranyl diphosphate synthase	GPSS	2.5.1.1	5	18.14	6.35	9.33	21.13	3	31.29	13.66	12.14	29.57		
	Geraniol 10-hydroxylase	G10H	1.14.14.83	11	69.38	156.81	124.65	18.33	0	-	-	-	-		
	Cytochrome P450 reductase	CPR	1.6.2.4	5	39.03	80.27	17.93	46.62	15	32.62	34.61	17.62	47.10		
	10-Hydroxygeraniol oxidoreductase	10-HGO	1.1.1.324	0	-	-	-	-	2	134.67	472.52	539.78	321.00		
	Iridoid synthase	IS	1.3.1.99	1	0.15	0.04	10.98	0.55	1	0.57	5.33	3.32	1.09		
	Iridoid	7-Deoxyloganic acid glucosyltransferase	7-DLGT	2.4.1.323	8	2.73	3.74	6.71	7.06	5	0.68	1.32	2.19	2.13	
7-Deoxyloganic acid hydroxylase		7-DLH	-	1	0.99	4.69	14.00	0.56	0	-	-	-	-		
Secologanin synthase		SLS	1.3.3.9	4	633.93	35.77	3.68	500.77	7	560.74	1.57	0.14	345.94		
Strictosidine synthase		STR	4.3.3.2	4	0.99	0.76	1.15	0.43	1	0.00	0.00	0.00	3.85		
Geranylgeryl pyrophosphate synthase		GGPPS	1.3.1.74	8	5.93	7.89	8.75	9.70	3	9.43	21.42	28.14	15.62		



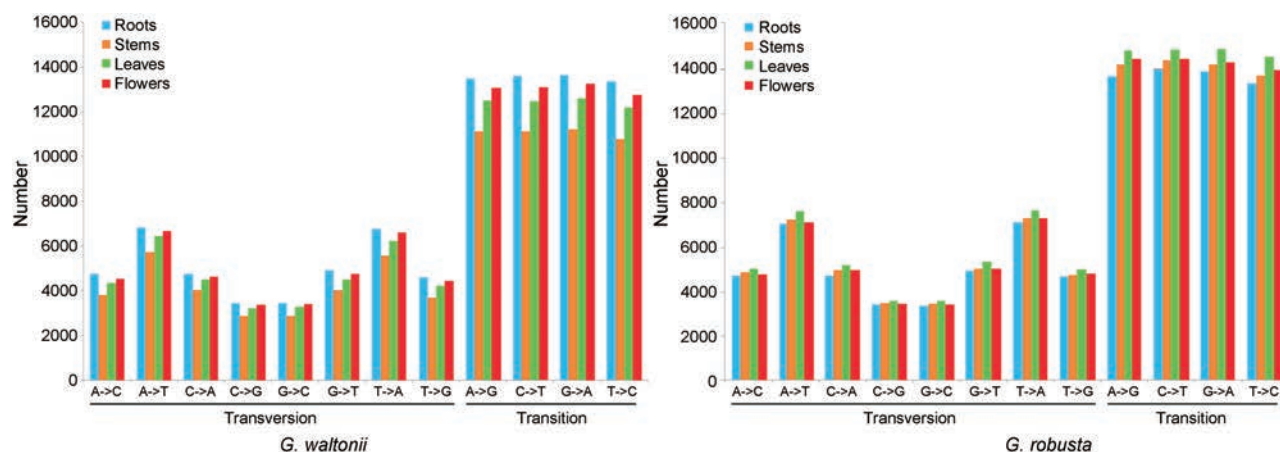


Figure 9 SNP types of transcriptomic unigene

得到注释的部分, 是否为长度较短未得到有效注释, 非编码序列或为新的功能未知基因<sup>[22]</sup>, 有待进一步研究。

根据 KEGG 数据库, 对长梗秦艽、粗壮秦艽转录组数据进行代谢途径分析, 注释到萜类、黄酮类、苯丙素类和生物碱类等 20 个次生代谢通路基因共 315 条和 340 条。作为秦艽组药用植物标志性成分及活性成分之一, 环烯醚萜苷类代谢过程主要经过中间体生成、萜类的合成和修饰等步骤。长梗秦艽、粗壮秦艽转录组中分别有 80 条、57 条 unigene 参与编码环烯醚萜苷类合成途径中的 24 个关键酶, 包括 MVA 和 MEP 途径中所有的基因 (除了粗壮秦艽的 CMK 基因) 及大部分已知的环烯醚萜苷骨架合成与修饰基因。未注释到的基因其原因可能为: ① 与其生长发育状态有关<sup>[22]</sup>; ② 居群生存环境胁迫; ③ 物种间差异等。就不同器官的表达情况来看, 各基因在根、茎、叶、花等不同部位的表达量具有较明显的差异。综合所有 24 个关键酶相关 unigene 的平均 TPM 值, 在两个物种中, 均为根最高, 其次为花, 而茎和叶部位的值相对较低。值得注意的是裂环马钱子苷合成酶 (SLS) 基因, 两物种根和花中基于 TPM 值的表达量均最高, 达 345.94~633.93。

中药秦艽及龙胆, 基原植物多分布于青藏高原, 生存环境严酷。由于种种原因, 资源状况不容乐观, 竖龙胆 *Gentiana rigescens* Franch. ex Hemsl.、粗茎秦艽 *Gentiana crassicaulis* Duthie ex Burk. 等相关 8 种龙胆属植物均已被列入《国家重点保护野生药材物种名录》之中<sup>[23]</sup>。其药用根部, 药材采挖过程中地上茎、叶及花部则弃之不用。如何结合不同部位次生产物合成及积累特点进行新药用部位的开发研究, 就地保护资源, 合理利用资源, 值得深入探讨。下一步将选择相关关键酶基因, 基于 qRT-PCR (quantitative real-time PCR) 进行系统的定量表达分析, 并开展不同部位环烯醚萜苷类合成和积累等相关基础性研究工作。

基于转录组测序的 SSR 和 SNP 等分子标记已广泛应用于药用植物物种鉴定、遗传多样性分析等方面<sup>[24,25]</sup>。本研究结果显示: 长梗秦艽和粗壮秦艽转录组中具有丰度较高的 SSR, 其中单核苷酸占大多数, 其次主要为三核苷酸和二核苷酸, 式样与麻花秦艽及川西獐牙菜等龙胆科植物相似<sup>[26,27]</sup>。同时, 以各物种最终拼接的 unigene 为 reference, 分别在长梗秦艽和粗壮秦艽的根、茎、叶和花的数据中检测得到大量的 SNP 和 InDel 位点信息, 可为新分子标记的开发提供丰富的数据资料。

## References

- [1] Yutok Yonten Gonpo. The Four Medical Tantras (四部医典) [M]. Shanghai: Shanghai Science & Technology Press, 1987: 66.
- [2] Dimaer Danzeng Pengcuo. Jing Zhu Ben Cao (晶珠本草) [M]. Shanghai: Shanghai Science & Technology Press, 1986: 105-106, 119.
- [3] Yang YC. Tibetan Medicine (藏药志) [M]. Xining: Qinghai People Press, 1991: 9-12, 213-218.
- [4] Zhao ZL, Dorje Gaawe, Wang ZT. Identification of medicinal plants used as Tibetan traditional medicine Jie-Ji [J]. J Ethnopharmacol, 2010, 132: 122-126.
- [5] Ni LH, Zhao ZL. A morphometric comparison of three closely related species of *Gentiana* (Gentianaceae), endemic to the region of the Qinghai-Tibet Plateau [J]. Botany, 2018, 96: 209-215.
- [6] Ni LH, Zhao ZL, Xu HX, et al. Chloroplast genome structures in *Gentiana* (Gentianaceae), based on three medicinal alpine plants used in Tibetan herbal medicine [J]. Curr Genet, 2017, 63: 241-252.
- [7] Ni LH, Zhao ZL, Xu HX, et al. The complete chloroplast genome of *Gentiana straminea*, (Gentianaceae), an endemic species to the Sino-Himalayan subregion [J]. Gene, 2016, 577: 281-288.
- [8] Xiong B, Zhao ZL, Ni LH, et al. DNA-based identification of

- Gentiana robusta* and related species [J]. *China J Chin Mater Med* (中国中药杂志), 2015, 40: 4680-4685.
- [9] Ni LH, Zhao ZL, Xiong B, et al. A strategy for identifying six species of Sect. *Cruciata* (*Gentiana*) in Gansu using DNA barcode sequences [J]. *Acta Pharm Sin* (药学报), 2016, 51: 821-827.
- [10] Wei SJ, Zhao ZL, Ni LH, et al. Taxonomic identification of Tibetan herb Bawo Sebo and its chloroplast genome structure [J]. *Acta Pharm Sin* (药学报), 2018, 53: 1009-1015.
- [11] Zheng LS, Lu XQ. Advances in the research of iridoids [J]. *Nat Prod Res Dev* (天然产物研究与开发), 2009, 21: 702-711, 725.
- [12] Chinese Pharmacopoeia Committee. Pharmacopoeia of the People's Republic of China. Part 1 (中华人民共和国药典) [S]. Beijing: China Medical Science Press, 2015: 96, 134-135, 196-197, 270-271.
- [13] Ozsolak F, Milos PM. RNA sequencing: advances, challenges and opportunities [J]. *Nat Rev Genet*, 2011, 12: 87-98.
- [14] Wang YL, Huang LQ, Yuan Y, et al. Research advances on analysis of medicinal plants transcriptome [J]. *China J Chin Mater Med* (中国中药杂志), 2015, 40: 2055-2061.
- [15] Grabherr MG, Haas BJ, Yassour M, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome [J]. *Nat Biotechnol*, 2011, 29: 644-652.
- [16] Thiel T, Michalek W, Varshney RK, et al. Exploiting EST databases for the development and characterization of gene-derived SSR - markers in barley (*Hordeum vulgare* L.) [J]. *Theor Appl Genet*, 2003, 106: 411-422.
- [17] Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data [J]. *Bioinformatics*, 2011, 27: 2987-2993.
- [18] Contin A, Heijden RVD, Lefeber AWM, et al. The iridoid glucoside secologanin is derived from the novel triose phosphate/pyruvate pathway in a *Catharanthus roseus* cell culture [J]. *Febs Lett*, 1998, 434: 413-416.
- [19] Yang R, Fang L, Li J, et al. Research progress on biosynthetic pathways and related enzymes of iridoid glycosides [J]. *Chin Tradit Herb Drugs* (中草药), 2018, 49: 2482-2488.
- [20] Wu XY, Liu XL. Progress of biosynthetic pathway and the key enzyme genes of iridoids [J]. *Chin J Ethnomed Ethnopharm* (中国民族民间医药), 2017, 26: 44-48.
- [21] Wagner GP, Kin K, Lynch VJ. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples [J]. *Theory Biosci*, 2012, 131: 281-285.
- [22] Li H, Zhang N, Li YM, et al. High-throughput transcriptomic sequencing of *Rheum palmatum* L. seedlings and elucidation of genes in anthraquinone biosynthesis [J]. *Acta Pharm Sin* (药学报), 2018, 53: 1908-1917.
- [23] Zhou XJ, Xu HF, Shun QS. Resource Science of Chinese Medicinal Materials (中药资源学) [M]. Shanghai: Shanghai Scientific and Technological Literature Publishing House, 2007: 370.
- [24] Jiang C, Yuan Y, Liu GM, et al. EST-SSR identification of *Lonicera japonica* Thunb. [J]. *Acta Pharm Sin* (药学报), 2012, 47: 803-810.
- [25] Kim HJ, Jung J, Kim MS, et al. Molecular marker development and genetic diversity exploration by RNA-Seq in *Platycodon Grandiflorum* [J]. *Genome*, 2015, 58: 1-11.
- [26] Zhou D, Gao S, Wang H, et al. *De novo* sequencing transcriptome of endemic *Gentiana straminea* (Gentianaceae) to identify genes involved in the biosynthesis of active ingredients [J]. *Gene*, 2015, 575: 160-170.
- [27] Liu Y, Yue CJ, Wang Y, et al. Data mining of simple sequence repeats in transcriptome sequences of Tibetan medicinal plant Zangyinchen *Swertia mussotii* [J]. *China J Chin Mater Med* (中国中药杂志), 2015, 40: 2068-2076.