

DOI: 10.11931/guihaia.gxzw202207050

韩霜, 徐浩, 余静雅, 等. 2023. 藏茵陈基源植物皱边喉毛花的全长转录组信息分析 [J]. 广西植物, 43(7): 1335–1346.
HAN S, XU H, YU JY, et al., 2023. Full-length transcriptome information for Tibetan medicine “Zangyinchen” of original plant *Comastoma polycladum* [J]. *Guihaia*, 43(7): 1335–1346.



藏茵陈基源植物皱边喉毛花的全长转录组信息分析

韩霜^{1,2}, 徐浩^{1,2}, 余静雅^{1,2}, 韩赞^{1,2}, 张发起^{1*}

(1. 中国科学院西北高原生物研究所 高原生物适应与进化重点实验室, 西宁 810001; 2. 中国科学院大学 生命科学学院, 北京 100039)

摘要: 皱边喉毛花为藏药藏茵陈基源植物之一, 其包含丰富的药用成分。为进一步了解皱边喉毛花转录组, 丰富其基因注释、代谢通路等遗传信息, 该研究利用 PacBio 测序平台对皱边喉毛花叶片进行全长转录组测序。结果表明: (1) 全长转录组测序共获得 17 Gb 的高质量数据, 对 795 698 个环形一致性序列 (CCS) 序列进行聚类 and 去冗余, 最终获得 87 814 条高质量的全长转录本。(2) 与 7 个数据库比对后, 共有 277 451 条转录本注释成功, 其中注释到 NR 数据库的转录本最多, 有 39 214 条。26 396 条转录本成功注释到 KOG 数据库中, 共有 26 个子类。39 104 条转录本注释到 KEGG 数据库中, 涉及 6 个主要通路和 40 个子通路。39 102 条转录本注释到 GO 数据库中, 按分子功能、生物学过程和细胞成分 3 大类对注释成功的转录本进行分类。(3) SSR 分析共鉴定到 22 861 个 SSR, 其中单碱基重复最为丰富; 共检测到 1 874 个转录因子和 15 166 个长非编码 RNA (lncRNA), 而注释到转录本最多的转录因子家族是 C3H。(4) 筛选出 55 条与单萜类及黄酮类化合物合成相关的转录本。该研究结果丰富了皱边喉毛花的转录组信息, 为进一步筛选皱边喉毛花药用成分合成相关的关键基因提供了重要的遗传资源。

关键词: 皱边喉毛花, 全长转录组, 代谢通路, 转录因子, 长非编码 RNA

中图分类号: Q943 文献标识码: A 文章编号: 1000-3142(2023)07-1335-12

Full-length transcriptome information for Tibetan medicine “Zangyinchen” of original plant *Comastoma polycladum*

HAN Shuang^{1,2}, XU Hao^{1,2}, YU Jingya^{1,2}, HAN Yun^{1,2}, ZHANG Faqi^{1*}

(1. Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining 810001, China; 2. College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100039, China)

Abstract: *Comastoma polycladum* is one of the original plant of Tibetan medicine “Zangyinchen”, which contains abundant medical components. To further know the transcriptome of *C. polycladum* and enrich its genetic information of gene annotation and metabolic pathway, the Pacbio sequencing platform was used to perform full-length transcriptome sequencing of *C. polycladum* leaves. The results were as follows: (1) A total of 17 Gb of sequencing data was collected,

收稿日期: 2022-11-21

基金项目: 第二次青藏高原科学考察研究项目 (2019QZKK05020102); 青海省科技国际合作专项 (2021-HZ-807)。

第一作者: 韩霜 (1998-), 硕士研究生, 主要从事高山植物多样性研究, (E-mail) hanshuang@nwipb.cas.cn。

*通信作者: 张发起, 博士, 研究员, 研究方向为高山植物多样性, (E-mail) fqzhang@nwipb.cas.cn。

and 87 814 high-quality full-length transcripts were obtained by clustering and de-redundancy of 795 698 circular consistency sequences (CCSs) sequences. (2) Comparing with seven databases, 277 451 transcripts were annotated successfully, and in NR database with 39 214 transcripts annotated the most transcripts. A total of 26 396 transcripts were annotated to the KOG database, with 26 subcategories, and a total of 39 104 transcripts with six major pathways and 40 secondary pathways to the KEGG database. A total of 39 102 transcripts were annotated to the GO database, which were divided into three major categories: molecular function, biological process and cellular component. (3) SSR analysis yielded 22 861 SSRs, with single-base repeats being the most abundant. A total of 1 874 transcription factors and 15 166 long non-coding RNAs (LncRNAs) were identified, and the C3H transcription factor family had the most annotated transcripts. (4) A total of 55 transcripts involved in the synthesis of monoterpenes and flavonoids were screened out. These results enrich the transcriptome information of *C. polycladum*, and provide significant genetic resources for further screening of candidate genes related to the synthesis of medicinal components of *C. polycladum*.

Key words: *Comastoma polycladum*, full-length transcriptome, metabolic pathway, transcription factor, LncRNA

藏茵陈是青藏高原藏药八珍之一,龙胆科植物是藏茵陈入药源植物中的主要植物,多以川西獐牙菜、湿生扁蕾、椭圆叶花锚和喉毛花属植物入药,常用于热症、肝胆病及血液病等疾病的治疗(唐丽等,2007)。近年来的研究表明,这些基源植物包含丰富的药用成分,主要为环烯醚萜、黄酮类化合物,在保肝、抗氧化、抗病毒等方面具有显著效果(延玺等,2008;董天骄等,2011;杨青松等,2013)。龙胆科(Gentianaceae)喉毛花属(*Comastoma*)植物是藏茵陈基源植物之一(钟国跃等,2010),对喉毛花属植物的研究目前主要集中在细胞学、胚胎学、生态学、系统发育研究及天然产物学上(刘建全和何廷农,1996;张婵等,2014;刘小翠等,2016;Zhang et al., 2021;刘真等,2021)。刘真等(2021)在长梗喉毛花的化学成分研究中发现 25 个化合物,其抗炎活性较高并对人体癌细胞株具有抑制作用。乔涌起等(2012)在长梗喉毛花植物中分离得到正丁醇化学成分,为进一步深入研究其化学成分奠定基础。然而,有关喉毛花属植物的基因注释信息尚未见报道,限制了对次级代谢产物合成相关代谢通路及功能基因的研究。因此,需要利用测序技术丰富喉毛花植物的转录组遗传信息。

随着测序技术的发展,越来越多的学者将高通量测序技术应用到植物转录组研究(Wang et al., 2016;Zhu et al., 2017;朱兴正等,2018)。二代测序读长的限制导致所拼接得到的转录本不够完整,而三代测序技术正好弥补了这一缺点,其能够完成长读长测序,测序过程无需打断,严格执行

RNA 样品提取与检测、建库及测序等环节的工作,最终得到高质量的全长转录本信息(王瑞娴和李川,2019;张子敬等,2020)。对没有参考基因组的植物而言,全长转录组(full-length transcriptome)测序为其研究提供了可能,解决了转录本拼接较短、信息不完整的难题(赵陆滢等,2019)。因此,三代测序技术成为深入挖掘基因组数据的有效手段之一(赵陆滢等,2019)。近年来,有许多学者研究了青藏高原地区药用植物的全长转录组。在这些研究案例中,对老芒麦(*Elymus sibiricus*)的转录组解析成功并挖掘到其落粒相关候选基因,为筛选低落粒老芒麦新品种提供了参考(张俊超,2020)。丹参(*Salvia miltiorrhiza*)的全长转录组揭示了丹参酮二萜类化合物的生物合成的相关基因(Xu et al., 2015)。蒙古黄芪(*Astragalus membranaceus* var. *mongholicus*)全长转录组解析了次生代谢产物生物合成的相关基因(Li et al., 2017)。这些研究案例说明全长转录组对药用植物关键基因的挖掘具有显著优势,为进一步研究药用植物的功能基因提供了新的思路和参考。

喉毛花属的皱边喉毛花(*Comastoma polycladum*)为青藏高原特有植物(Ho & Liu, 2001, 2015)。目前对皱边喉毛花的研究主要集中在系统发育研究上。为进一步了解喉毛花属下物种植物体内的次级代谢产物,应对相关转录组进行深入研究,本研究以皱边喉毛花为对象,基于 PacBio 测序平台对其全长转录组进行测序,获取的数据用于功能注释、可变剪切分析、SSR 分析、转录因子分析及长非编码 RNA 等分析。通过与公共数

数据库比对,筛选药用相关成分合成相关的代谢通路和转录本。全长转录组能够为皱边喉毛花药用成分合成相关的关键基因的筛选提供重要的遗传资源。

1 材料与方法

1.1 试验材料

新鲜幼叶采集于海南藏族自治州共和县(地理坐标为 100°53'58.44" E、36°22'19.00" N,海拔为 3 518 m),采集后迅速置于液氮罐中保存,后将其转移至-80 °C 的超低温冰箱中,用于后续 RNA 提取。凭证标本(Zhang2018026)存放于中国科学院西北高原生物研究所青藏高原生物标本馆(HNWP)。

1.2 方法

1.2.1 RNA 提取和文库构建 采用 Total RNA Extractor(Trizol)试剂法(Connolly et al., 2006)提取皱边喉毛花的总 RNA,琼脂糖凝胶电泳检测 RNA 降解程度及污染情况,并评估其质量和完整性。检测合格的 RNA 样品用于构建皱边喉毛花全长转录组测序文库。具体操作如下:在反转录酶的作用下,以 Oligo DT 为引物、目标 mRNA 为模板,进行反转录,通过低循环 PCR 扩增全长 cDNA,利用 NEBNext End repair/dA-tailing Module 末端修读及加 poly(A)尾,ONT SQKLSK109 试剂盒及 NEBNext Quick Ligation Module 用于测序接头的连接。建好的文库采用 PromethION(Oxford Nanopore Technologies 公司,英国)测序平台进行测序。

1.2.2 数据处理 测序完成后对原始数据进行过滤,去除接头以及低质量的 reads。采用软件 SMRTlink v8.0 (<https://www.pacb.com/support/software-downloads>)进行过滤和处理。参数设置: -minLength 50(最小长度为 50 bp), -maxLength 15 000(最大长度为 15 000 bp), -minPasses 1(最小的 fullpass 数为 1)。利用 subread.bam 文件得到环形一致性序列(circular consistency sequence, CCS),对其进行分类,搜寻并聚类 FLNC 序列,得到 consensus 序列。利用 Arrow 软件对得到的 consensus 序列进行校正,获得高质量的全长优化序列(polished consensus)用于后续分析,最终统计得到有效数据。为提高数据的准确性,利用 LoRDEC 软件(Salmela & Rivals, 2014)对转录本进

行校正,生成校正序列(corrected consensus)。利用 CD-HIT 软件(Fu et al., 2012)对校正后的转录本进行冗余分析。

1.2.3 全长转录组序列分析 对去冗余后的序列进行基因功能注释,所使用的数据库包括非冗余蛋白数据库(Non-Redundant Protein Database, NR, Deng et al., 2006)、蛋白质家族域数据库(Protein Families Database, Pfam, Finn et al., 2014)、蛋白质真核同源数据库(Eukaryotic Orthologous Groups, KOG, Koonin et al., 2004)、蛋白质原核同源数据库(Cluster of Orthologous Groups of Proteins, COG, Tatusov et al., 2003)、东京基因与基因组百科全书(Kyoto Encyclopedia of Genes and Genomes, KEGG, Minoru et al., 2004)、基因本体论数据库(Gene Ontology, GO, Ashburner et al., 2000)、核酸序列数据库(Nucleotide Sequence Database, NT)和 SwissProt 数据库(a manually annotated and reviewed protein sequence database)等,以此获得更全面的基因功能信息。

1.2.4 全长转录组结构分析 利用 iTAK v 1.7a 软件(参数设置: -f 3F)对皱边喉毛花进行转录因子(transcription factor, TF)预测(Zheng et al., 2016)。利用 MISA(MicroSATellite Identification Tool) v 1.0 软件检测简单重复序列标记(simple sequence repeats, SSR),设置单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸及六核苷酸的最少重复次数分别为 10、6、5、5、5、5,其余参数默认(Beier et al., 2017)。利用 CNCI v 2(Coding-Non-Coding Index, Sun et al., 2013)、plek v 1.2(Predictor of Long Non-coding RNAs and mRNAs Based on k-mer Scheme, Li et al., 2014)、CPC2 v 0.1(Coding Potential Calculator 2, Kang et al., 2017)软件以及 Pfam 数据库(Finn et al., 2014)对 PacBio 测序数据进行编码潜能预测(参数设为默认),获得的长非编码 RNA(long non-coding RNA, LncRNA)用于后续分析。

2 结果与分析

2.1 全长转录组测序及组装

经 SMRT(Single-Molecule Real-Time)测序共获得 17 Gb 的原始数据。对其过滤后获得 17 315 066 个 subreads。基于 subread.bam 文件获得 795 698 个 CCS 序列, N50 长度为 2 143 bp,最

大长度为 16 860 bp,最小长度为 60 bp,平均长度为 2 337 bp(表 1)。经分类获得 695 698 条 FLNC 序列,N50 长度为 2 055 bp,最大长度为 12 952 bp,最小长度为 52 bp。对其聚类后获得 87 814 条 consensus 序列,最大长度为 8 269 bp,最小长度为 338 bp,N50 长度为 2 219 bp。对 consensus 序列进行校正后,获得 87 814 条校正序列,N50 长度为 2 221 bp。我们对冗余前后的序列长度频数分布情况进行了统计(图 1)。

表 1 SMRT 测序数据统计
Table 1 Statistics of sequencing data by SMRT

类别 Category	总数 Total account	最小 长度 Min length (bp)	最大 长度 Max length (bp)	平均 长度 Mean length (bp)	N50 长度 N50 length (bp)
Subread	17 315 066	—	—	1 620	1 505
环形一致性序列 CCS	795 698	60	16 860	2 337	2 143
FLNC	695 698	52	12 952	1 774	2 055
优化序列 Polished sequence	87 814	338	8 269	2 233	2 219
校正序列 Corrected sequence	87 814	338	8 269	2 233	2 221

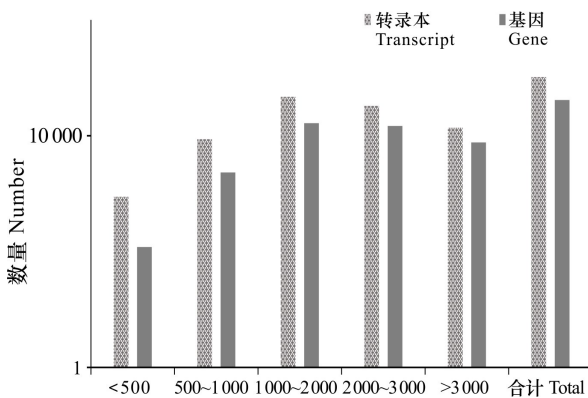


图 1 皱边喉毛花全长转录组 Transcripts 和 Unigenes 长度分布图

Fig. 1 Full-length transcriptome of distribution of Transcripts and Unigenes for *Comastoma polycladum*

2.2 全长转录组功能注释

共有 277 451 条转录本成功注释到 7 个数据库中,其中注释到 NR 数据库的转录本最多,有

39 214 条,注释到 KOG 数据库的转录本最少,有 26 396 条(图 2)。26 722、39 102 条转录本分别注释到 GO 数据库和 KEGG 数据库中。39 592 条转录本注释到至少一个数据库中,16 273 条转录本注释到所有数据库中。从不同数据库中选择 5 个常用的数据库进行韦恩图绘制(图 3)。

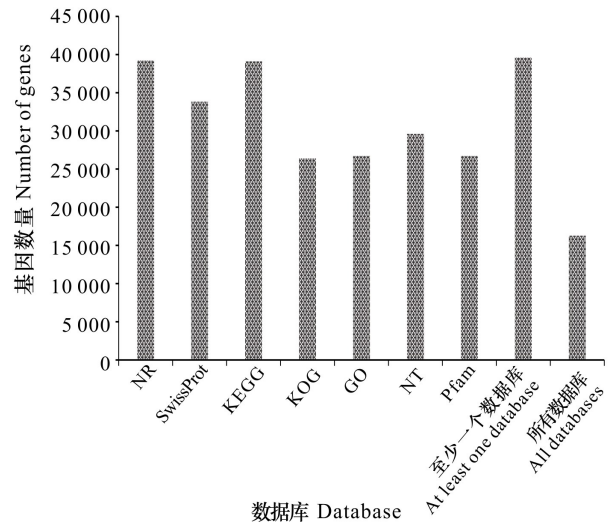


图 2 转录本注释结果

Fig. 2 Annotation results of transcripts

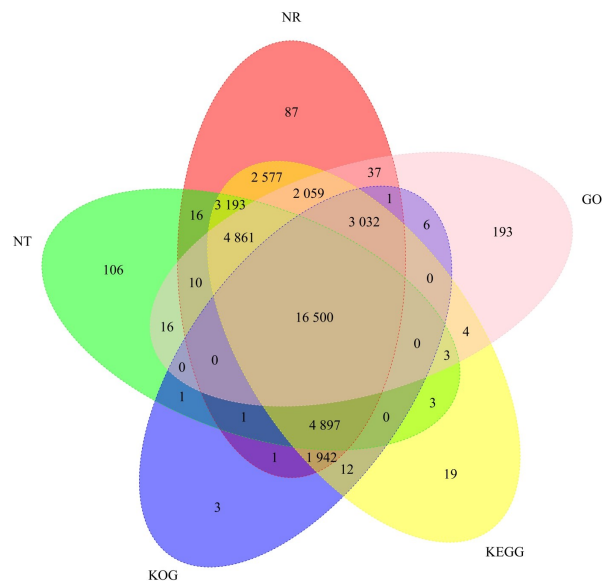
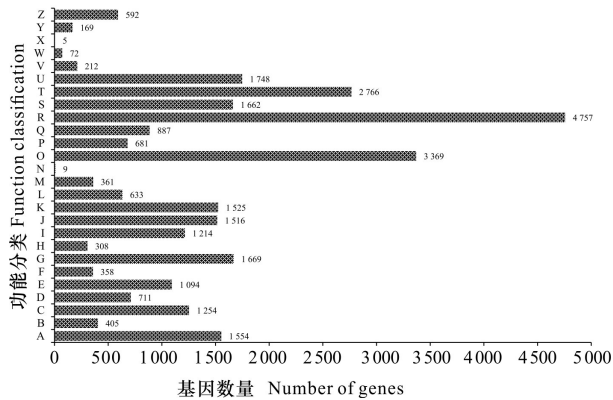


图 3 转录本功能注释韦恩图

Fig. 3 Venn diagram of function annotated transcripts

2.2.1 KOG 注释 与 KOG 数据库比对后,共有 29 531 条基因被注释成功。按 KOG 分类可分为 26

个类型(图4)。其中,注释到只有一般功能预测(4 757 条)、翻译后修饰、蛋白转运(3 369 条)和信号传递机制(2 766 条)的基因最多。然而,细胞活性(9 条)和未命名蛋白(5 条)注释到的基因最少。



A. RNA 加工和修饰; B. 染色体结构和动力学; C. 能源生产与转化; D. 细胞周期调控、细胞分裂、染色体分离; E. 氨基酸转运和代谢; F. 核酸转运和代谢; G. 碳水化合物转运和代谢; H. 辅酶转运和代谢; I. 脂类转运和代谢; J. 翻译、核糖体结构和生物发生; K. 转录; L. 复制、重组和修饰; M. 细胞壁/细胞膜生物发生; N. 细胞活性; O. 翻译后修饰、蛋白转运; P. 无机离子转运和代谢; Q. 次生代谢物合成、转运和代谢; R. 只有一般功能预测; S. 未知功能; T. 信号传递机制; U. 细胞间运输、分泌物和囊泡运动; V. 防御机制; W. 细胞外结构; X. 未命名蛋白; Y. 核结构; Z. 细胞骨架。

A. RNA processing and modification; B. Chromatin structure and dynamics; C. Energy production and conversion; D. Cell cycle control, cell division, chromosome partitioning; E. Amino acid transport and metabolism; F. Nucleotide transport and metabolism; G. Carbohydrate transport and metabolism; H. Coenzyme transport and metabolism; I. Lipid transport and metabolism; J. Translation, ribosomal structure and biogenesis; K. Transcription; L. Replication, recombination and repair; M. Cell wall/membrane/envelope biogenesis; N. Cell motility; O. Posttranslational modification, protein turnover, chaperones; P. Inorganic ion transport and metabolism; Q. Secondary metabolites biosynthesis, transport and catabolism; R. General function prediction only; S. Function unknown; T. Signal transduction mechanism; U. Intracellular trafficking, secretion and vesicular transport; V. Defense mechanism; W. Extracellular structure; X. Unnamed protein; Y. Nuclear structure; Z. Cytoskeleton.

图4 转录本 KOG 注释结果

Fig. 4 KOG annotation results of transcripts

2.2.2 GO 注释 GO 是描述基因功能的一套分类系统,可全面描述生物体中基因和基因产物的属性(Blake & Harris, 2008)。共有39 102 条转录本注释

到了 GO 数据库,根据注释结果对得到的转录本进行分类(图5)。注释到生物过程、细胞成分、分子功能3 大类,分别有 25 个、18 个、11 个子类,共 54 个子类。然而,注释到细胞杀死(2 条)和行为(1 条)的基因最少。在细胞成分中,注释到细胞和细胞部分的基因最多(均为5 252 个),而突触(5 条)、突触部分(5 条)和细胞连接(4 条)子类中涉及的基因最少;在分子功能中,注释到结合(16 037 条)和催化活性(12 400 条)的基因最多,而金属伴活动分子功能调节器(2 条)子类中涉及的基因最少;在生物过程中,注释到代谢过程(12 848 条)和细胞过程(11 558 条)的基因最多,而细胞杀死(2 条)和行为(1 条)子类中涉及的基因最少。

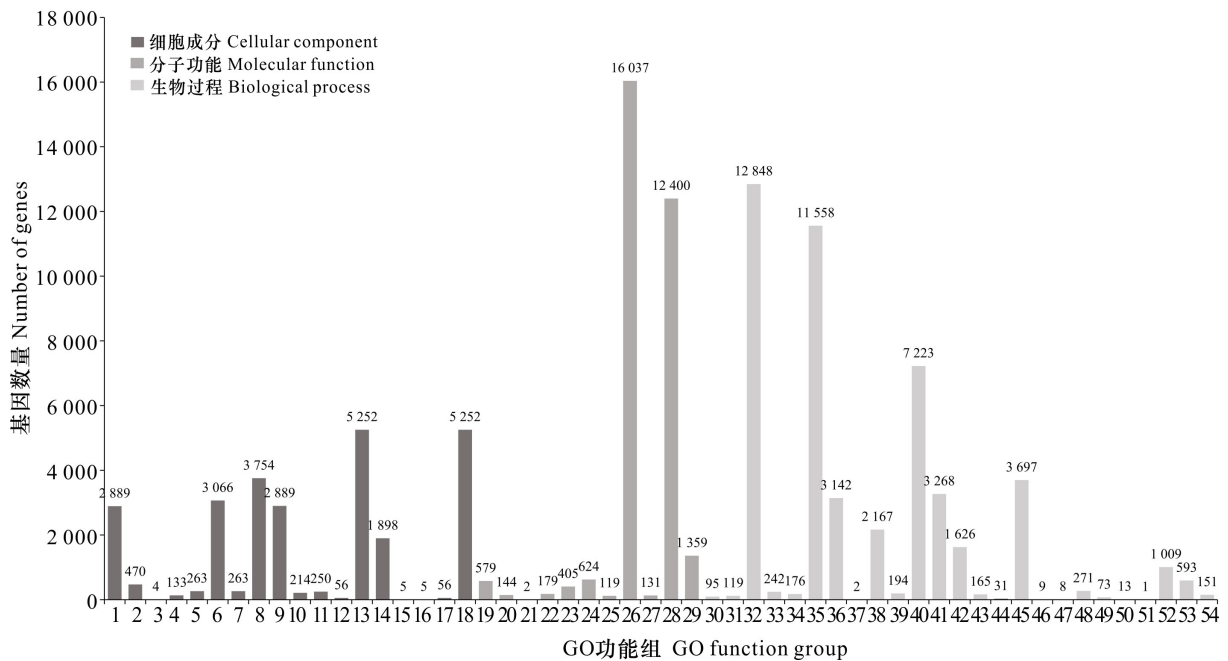
2.2.3 KEGG 注释 根据 KO 与 Pathway 的关联性进行 KEGG 代谢通路分类。与 KEGG 数据库比对应后,成功注释的基因有 62 457 条。这些注释基因被分到 6 个主要通路(代谢、遗传信息处理、细胞过程、环境信息处理、组织系统和人类疾病)和 40 个子通路中(表2)。其中,涉及基因最多的主通路为代谢(128 条),其次为组织系统(77 条)、人类疾病(68 条)和环境信息处理(36 条),最少为细胞过程(25 条)和遗传信息处理(22 条)。涉及基因最多的子通路为信号转导(3 283 条),其次为碳代谢(2 468 条)和翻译(2 460 条),最少为信号分子和互作作用(1 条)。

2.3 药用相关的代谢通路

龙胆科植物包含环烯醚萜、黄酮类及三萜类化合物等药效成分(杨青松等,2013)。根据 KEGG 转录本注释结果,统计及分析与药效成分相关的次级代谢通路(表3)。其中包括单萜类生物合成(13 条)、倍半萜类和三萜类生物合成(27 条)、类黄酮生物合成(36 条)、黄酮和黄酮醇生物合成(4 条)。统计这些代谢通路中可能与环烯醚萜、黄酮等药效成分合成相关的转录本。

2.4 SSR 分析

对皱边喉毛花的全长转录组进行 SSR 分析后,共检测到 22 861 个 SSR 信息位点。共获得 6 种 SSR 重复类型,其中单碱基重复类型(Mono-nucleotide, Mono-, 13 750 个)最丰富,其次为三碱基重复类型(Tri-nucleotide, Tri-, 5 811 个)、二碱基重复类型(Di-nucleotide, Di-, 2 807 个)、四碱基重复类型(Tetra-nucleotide, Tetra-, 573 个)和六碱基重复类型(Hexa-nucleotide, Hexa-, 199 个),而五碱



1. 膜部分; 2. 膜封闭腔; 3. 细胞连接; 4. 超分子纤维; 5. 病毒; 6. 膜; 7. 病毒部分; 8. 细胞器; 9. 大分子复合物; 10. 胞外区域部分; 11. 胞外区域; 12. 其他有机体; 13. 细胞; 14. 细胞器部分; 15. 突触; 16. 突触部分; 17. 其他有机体部分; 18. 细胞部分; 19. 结构分子活性; 20. 核酸结合转录因子活性; 21. 金属伴活动分子功能调节器; 22. 转录因子活性与蛋白质结合; 23. 分子功能调控; 24. 分子转导活性; 25. 抗氧化活性; 26. 结合; 27. 信号转导活性; 28. 催化活性; 29. 转运蛋白活性; 30. 生物黏附; 31. 生物过程负调控; 32. 代谢过程; 33. 多细胞生物过程; 34. 生殖过程; 35. 细胞过程; 36. 生物过程的调节; 37. 细胞杀死; 38. 生殖过程对刺激的反应; 39. 繁殖; 40. 单一生物体过程; 41. 生物调节; 42. 细胞成分组织或生物合成; 43. 生物过程正调控; 44. 生长; 45. 定位; 46. 生物相; 47. 节律过程; 48. 发育过程; 49. 免疫系统过程; 50. 解毒; 51. 行为; 52. 信号; 53. 多组织过程; 54. 转运。

1. Membrane part; 2. Membrane-enclosed lumen; 3. Cell junction; 4. Supramolecular fiber; 5. Virion; 6. Membrane; 7. Virion part; 8. Organelle; 9. Macromolecular complex; 10. Extracellular region part; 11. Extracellular region; 12. Other organism; 13. Cell; 14. Organelle part; 15. Synapse; 16. Synapse part; 17. Other organism part; 18. Cell part; 19. Structural molecule activity; 20. Molecular transducer activity; 21. Metallochaperone activity; 22. Transcription factor activity, protein binding; 23. Molecular function regulator; 24. Nucleic acid binding transcription factor activity; 25. Antioxidant activity; 26. Binding; 27. Signal transducer activity; 28. Catalytic activity; 29. Transporter activity; 30. Biological adhesion; 31. Negative regulation of biological process; 32. Metabolic process; 33. Multicellular organismal process; 34. Reproductive process; 35. Cellular process; 36. Regulation of biological process; 37. Cell killing; 38. Response of reproductive processes to stimulus; 39. Reproduction; 40. Single-organism process; 41. Biological regulation; 42. Cellular component organization or biogenesis; 43. Positive regulation of biological process; 44. Growth; 45. Localization; 46. Biological phase; 47. Rhythmic process; 48. Developmental process; 49. Immune system process; 50. Detoxification; 51. Behavior; 52. Signaling; 53. Multi-organism process; 54. Locomotion.

图5 转录本 GO 注释结果

Fig. 5 GO annotation results of transcripts

基重复类型 (Penta-nucleotide, Penta-, 178 个) 最少。统计 6 个重复类型中不同长度范围重复序列的分布情况, 结果显示 9~12 bp 长度的重复序列最丰富, 其次是 5~8 bp、13~16 bp、17~20 bp 长度的重复序列, 而 21~24 bp 长度的重复序列最少(图 6)。

2.5 转录因子分析

转录因子是一些表达的蛋白质分子, 能与基

因顺式作用元件专一性结合, 对基因转录进行调控(刘强等, 2000)。本研究预测结果显示共获得 1 874 个转录因子(图 7)。其中注释到转录本最多的转录因子家族是 C3H(120 个), 其次为 bZIP (Basic Leucine Zipper, 112 个)、bHLH (Basic/Helix-Loop-Helix, 110 个) 和 MYB-related (109 个), 而 TUB(31 个)、C2C2-GATA(30 个) 转录因子家族数量最少。

表 2 转录本 KEGG 注释结果
Table 2 KEGG annotation result of transcripts

KEGG 主通路 Main pathway of KEGG	KEGG 子通路 Sub-pathway of KEGG	转录本数量 Number of transcripts	KEGG 主通路 Main pathway of KEGG	KEGG 子通路 Sub-pathway of KEGG	转录本数量 Number of transcripts		
代谢 Metabolism	碳代谢 Carbohydrate metabolism	2 468	组织系统 Organismal system	环境适应 Environmental adaptation	401		
	脂肪代谢 Lipid metabolism	1 053		老化 Aging	384		
	辅助因子和维生素代谢 Metabolism of cofactors and vitamins	464		循环系统 Circulatory system	195		
	能量代谢 Energy metabolism	1 175		发育 Development	100		
	氨基酸代谢 Amino acid metabolism	1 440		内分泌系统 Endocrine system	1 053		
	核苷酸代谢 Nucleotide metabolism	475		免疫系统 Immune system	1 090		
	其他次生代谢物的生物合成 Biosynthesis of other secondary metabolites	409		神经系统 Nervous system	638		
	萜类化合物和聚酮类化合物的代谢 Metabolism of terpenoids and polyketides	376		感官系统 Sensory system	74		
	异种生物降解和新陈代谢 Xenobiotics biodegradation and metabolism	285		人类疾病 Human disease	药物抵抗 Drug resistance	192	
	其他氨基酸代谢 Metabolism of other amino acids	517			内分泌和代谢疾病 Endocrine and metabolic diseases	646	
	聚糖的生物合成和代谢 Glycan biosynthesis and metabolism	378			神经退行性疾病 Neurodegenerative diseases	903	
	全局和概览图 Global and overview maps	1 654			药物依赖 Substance dependence	177	
	环境信息处理 Environmental information processing	膜运输 Membrane transport			68	感染疾病: 细菌 Infectious diseases; Bacteria	898
		信号转导 Signal transduction			3 283	感染疾病: 病毒 Infectious diseases; Virus	1 485
信号分子和互作用 Signaling molecules and interaction		1	感染疾病: 寄生生物 Infectious diseases; Parasite		488		
遗传信息处理 Genetic information processing	翻译 Translation	2 460	癌症: 特殊类型 Cancers; specific types	1 386			
	折叠、分类和降解 Folding, sorting and degradation	1 442	心血管疾病 Cardiovascular diseases	235			
	复制和修复 Replication and repair	454	免疫疾病 Immune diseases	139			
细胞过程 Cellular process	细胞群落-原核生物 Cellular community - prokaryotes	482	运输和分解代谢 Transport and catabolism	1 390			
	细胞生长和死亡 Cell growth and death	974	细胞运动 Cell motility	116			

2.6 长非编码 RNA 分析

长链非编码 RNA (LncRNA) 不编码蛋白质, 利用 CNCI、CPC2、plek 和 Pfam 方法对其进行编码潜能预测。共检测到 15 166 个 LncRNA, 其中 CNCI、CPC2、plek 和 Pfam 分别检测到 2 276 个、5 295 个、6 655 个、10 480 个 LncRNA, 4 种软件检测到的共有 LncRNA 有 1 110 个。统计 4 种方法预测

的 LncRNA 数目绘制韦恩图(图 8)。

3 讨论与结论

皱边喉毛花是藏茵陈基源植物之一, 植物体内含有大量的药用成分, 但对其转录组的认识较浅, 基因功能相关研究的报道较少。为进一步了

表 3 萜类及黄酮化合物生物合成相关转录本

Table 3 Transcripts related to terpenoid and flavone biosynthesis

KEGG 代谢通路 Metabolism pathway of KEGG	通路编号 No. of pathway	转录本编号 Transcript IDs
单萜类生物合成 Monoterpenoid biosynthesis	Ko00902	transcript1977 transcript1735 transcript1709 transcript1648 transcript1440 transcript1281 transcript1465 transcript1357 transcript1230 transcript1262 transcript1229 transcript1134 transcript2753
		transcript3155 transcript2877 transcript2985 transcript2797 transcript2706 transcript2536 transcript2547 transcript2518 transcript2498 transcript2281 transcript2217 transcript2225 transcript2112 transcript2013 transcript1930 transcript1320 transcript1742 transcript3115
		transcript2107 transcript2120 transcript2025 transcript1751 transcript1657 transcript1725 transcript1652 transcript1597 transcript1594 transcript1571 transcript1526 transcript1396 transcript1845 transcript1458 transcript1409 transcript1287 transcript1242 transcript933 transcript2107 transcript2043
		transcript1842 transcript1821 transcript1759 transcript1845
黄酮和黄酮醇生 物合成 Flavone and flavonol biosynthesis	Ko00944	transcript1842 transcript1821 transcript1759 transcript1845

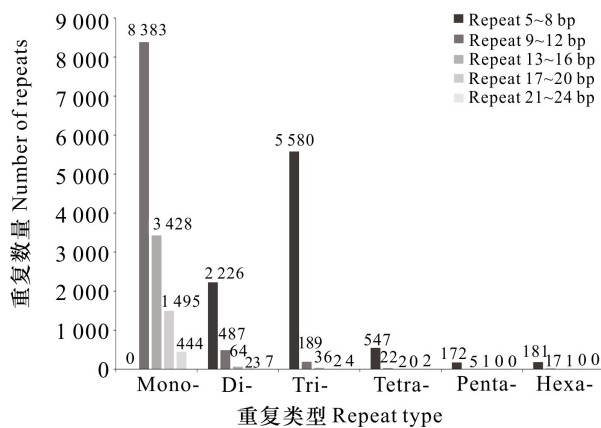


图 6 皱边喉毛花全长转录组 SSR 分析

Fig. 6 SSR analysis of full-length transcriptome of *Comastoma polycladum*

析全长转录组,获取更完整的转录本信息。通过测序共获得 17 315 066 条 Subreads,平均长度为 1 620 bp,N50 长度为 1 505 bp,说明全长转录组测序读长较长且连续性较高。为获得更准确且可靠的数据,对 Subread 进行聚类及校正后,得到 87 814 条高质量的全长转录本。N50 长度大于 1 000 bp(2 221 bp)说明其组装完整性较好,能够满足后续转录组信息分析的要求。

通过与 7 个公共数据库的比对,我们获得了大量有用的转录本注释信息,共有 277 451 条转录本成功注释到这些数据库中,其中注释到 NR、KEGG 数据库中的转录本数量最多。NR 数据库比对结果显示与皱边喉毛花比对率排名前十的物种为咖啡(*Coffea canephora*)、芝麻(*Sesamum indicum*)、牵牛(*Ipomoea nil*)等,表明皱边喉毛花与这些植物具有较高同源性。然而,能与皱边喉毛花比对上的同科植物较少,这反映出公共数据库中龙胆科植物的基因组信息十分匮乏。

藏茵陈基源植物包含丰富的环烯醚萜、黄酮及三萜类等药用成分(杨青松等,2013),本研究从对皱边喉毛花全长转录组的分析中检测到与其药用合成相关的 55 条转录本,包括生物碱、萜类、苯丙素类、黄酮类、糖苷类、醌类、聚酮类、有机酸及酚类等。黄酮作为具有多种生物活性的多酚类化合物,广泛存在于植物体内,在人体疾病治疗中起到显著效果(Hostetler et al.,2017)。萜类化合物在植物的生长和发育中发挥重要作用,常应用于食品、制药和化学工业中(Tholl,2015)。本研究分析得到多个黄酮类与萜类化合物合成相关的关键基因,这为我们今后开展皱边喉毛花关键基因克隆研究提供基础数据。

SSR 位点广泛分布于真核生物基因组(张楠等,2013)。本研究 SSR 分析结果显示,单碱基重复类型最为丰富,类似情况在其他药用植物中也有体现,如山萇蓉(*Anisodus tanguticus*)、凤丹(*Paeonia suffruticosa*)、罗布麻(*Apocynum venetum*)等(谢冬梅等,2017;张雨等,2020;赵雪艳等,2020)。除去对单碱基重复的统计,皱边喉毛花植物中三碱基重复类型最为丰富,这与药用植物甘葛藤(*Pueraria thomsonii*)和全萼秦艽(*Gentiana lhasica*)结果一致,而在红花这个物种中二碱基重复最为丰富,这种差异可能与 SSR 位点进化速率

解皱边喉毛花药用价值,我们开展次级代谢产物合成相关基因功能研究。利用 PacBio 测序平台解

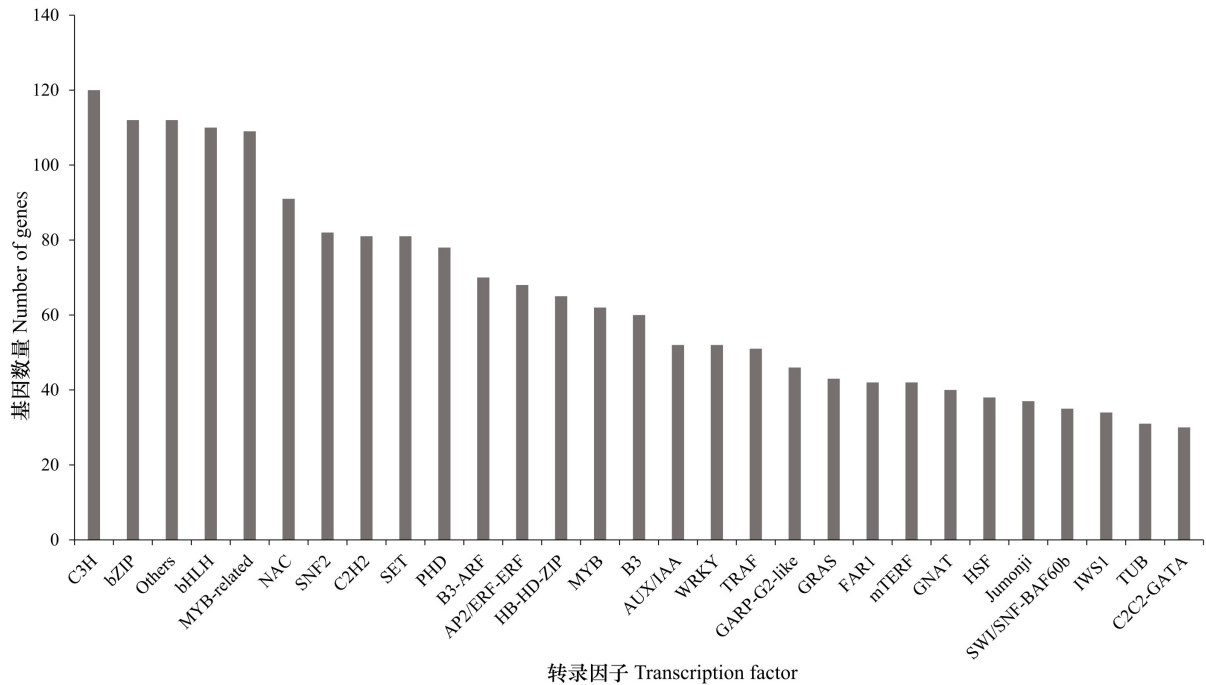


图 7 皱边喉毛花全长转录组转录因子家族分析

Fig. 7 Analysis of transcription factor family in full-length transcriptome of *Comastoma polycladum*

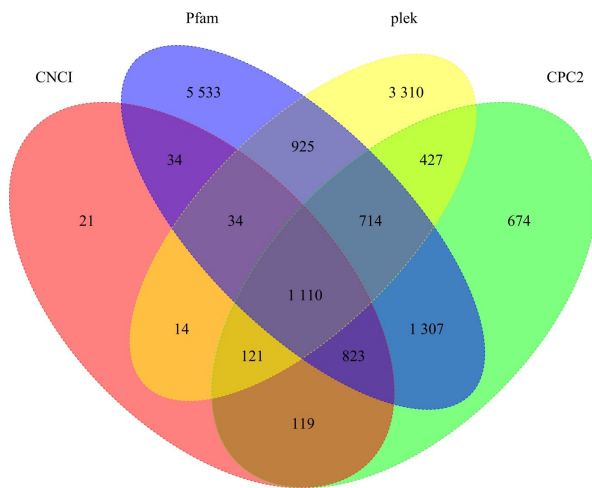


图 8 皱边喉毛花长非编码 RNA 分析结果

Fig. 8 LncRNA analysis result of *Comastoma polycladum*

不同、设定的检索参数以及样品来源有关 (Chen et al., 2018; 李延龙等, 2020; 梅瑜等, 2021)。转录因子是一类能够特异结合 DNA 且调节转录的蛋白质 (刘强等, 2000)。本研究鉴定出的 1 874 个转录因子中, C3H、bZIP、bHLH 及 MYB-related 转录

因子家族在皱边喉毛花中数量较多, 丰富了喉毛花属的转录因子信息。MYB-related 作为植物中最大的转录因子家族, 参与调控植物的生长发育、次生代谢及逆境胁迫等生物学过程 (陈清等, 2009)。在关于 MYB 转录因子家族的研究中, 前人已证实该基因家族能够提高植物的耐寒能力, 促进果实着色, 并在盐胁迫调控中发挥重要作用 (Allan et al., 2008; Dubos et al., 2010; Jaakola, 2013; 陈娜等, 2015)。bZIP 转录因子作为真核生物转录因子中分布最广、最保守的一类转录因子, 参与植物的生长发育、光信号转导、生物和非生物胁迫应答 (杨颖等, 2009)。已有研究证实 bZIP 转录因子家族能够增强拟南芥抗旱能力, 并参与其低温、高盐等胁迫的应答反应 (Choi et al., 2000; Kang et al., 2002)。bHLH 作为植物中最大的转录因子家族之一, 在细胞发育和细胞活性中发挥着重要作用, 并参与植物中次级代谢产物合成相关基因表达的调控, 如类黄酮、花青素等次级代谢物 (Gonzalez et al., 2008; 张全琪等, 2011; Zhao et al., 2013)。MYB-related 和 bZIP 转录因子家族与植物的非生物胁迫相关, 皱边喉毛花主要分布在青藏高原地区, 该地区气温低, 昼夜温差大, 这些转录因子家

族是否参与皱边喉毛花的响应胁迫应答反应, 还需进一步研究。LncRNAs 是一类长度超过 200 nt、能够调控基因的表达、参与多个生物学过程和通路的长非编码 RNA (Wan et al., 2019)。本研究共预测到 15 166 个 LncRNAs, 丰富了喉毛花属植物的长非编码 RNA 信息, 为后续进一步探索 LncRNA 在喉毛花属植物中的具体生物功能及作用机制提供了数据支持。

本研究利用 PacBio 测序平台对皱边喉毛花进行了测序及拼接。获得的大量测序数据用于功能注释、可变剪切分析、SSR 分析、转录因子分析及长非编码 RNA 预测等。通过与 KEGG 的比对, 筛选出药用相关成分合成相关的代谢通路和转录本, 为皱边喉毛花药用成分合成相关的关键基因的筛选提供重要的遗传资源。

参考文献:

- ALLAN AC, HELLENSE RP, LAING WA, 2008. MYB transcription factors that colour our fruit [J]. *Trends Plant Sci*, 13(3): 99-102.
- ASHBURNER M, BALL CA, BLAKE JA, et al., 2000. Gene Ontology: tool for the unification of biology [J]. *Nat Genet*, 25(1): 25-29.
- BEIER S, THIEL T, MÜNCH T, et al., 2017. MISA-web: a web server for microsatellite prediction [J]. *Bioinformatics*, 33(16): 2583-2585.
- BLAKE JA, HARRIS MA, 2008. The Gene Ontology (GO) project: structured vocabularies for molecular biology and their application to genome and expression analysis [J]. *Curr Protoc Bioinf*, 23(1): 721-729.
- CHEN J, TANG XH, REN CX, et al., 2018. Full-length transcriptome sequences and the identification of putative genes for flavonoid biosynthesis in safflower [J]. *BMC Genomics*, 19(1): 548.
- CHEN N, CHI XY, PAN LJ, et al., 2015. Advances in MYB transcription factors during salt-stress regulation in plants [J]. *Plant Physiol J*, 51(9): 1395-1399. [陈娜, 迟晓元, 潘丽娟, 等, 2015. MYB 转录因子在植物盐胁迫调控中的研究进展 [J]. *植物生理学报*, 51(9): 1395-1399.]
- CHEN Q, TANG HR, DONG XL, et al., 2009. Progress in the study of plant MYB transcription factors [J]. *Genomics Appl Biol*, 28(2): 365-372. [陈清, 汤浩茹, 董晓莉, 等, 2009. 植物 Myb 转录因子的研究进展 [J]. *基因组学与应用生物学*, 28(2): 365-372.]
- CHOI H, HONG J, HA J, et al., 2000. ABFs a family of ABA-responsive element binding factors [J]. *J Biol Chem*, 275(3): 1723-1730.
- CONNOLLY MA, CLAUSEN PA, LAZAR JG, 2006. Preparation of RNA from plant tissue using trizol [J]. *Cold Spring Harb Protoc*, 2006(1): pdb.prot4105. DOI: 10.1101/pdb.prot4105.
- DENG YY, LI JQ, WU SF, et al., 2006. Integrated NR database in protein annotation system and its localization [J]. *Comp Eng*, 32(5): 71-74.
- DONG TJ, CUI YL, TIAN JS, et al., 2011. Advances of studies on natural iridoids [J]. *Chin Trad Herb Drugs*, 42(1): 185-194. [董天骄, 崔元璐, 田俊生, 等, 2011. 天然环烯醚萜类化合物研究进展 [J]. *中草药*, 42(1): 185-194.]
- DUBOS C, STRACKE R, GROTEWOLD E, et al., 2010. MYB transcription factors in *Arabidopsis* [J]. *Trends Plant Sci*, 15(10): 573-581.
- FINN RD, BATEMAN A, CLEMENTS J, et al., 2014. Pfam: the protein families database [J]. *Nucl Acid Res*, 42(D1): D222-D230.
- FU LM, NIU BF, ZHU ZW, et al., 2012. CD-Hit: Accelerated for clustering the next-generation sequencing data [J]. *Bioinformatics*, 28(23): 3150-3152.
- GONZALEZ A, ZHAO MZ, LEAVITT JM, et al., 2008. Regulation of the anthocyanin biosynthetic pathway by the TTG1/bHLH/MYB transcriptional complex in *Arabidopsis* seedlings [J]. *Plant J*, 53(5): 814-827.
- HO TN, LIU SW, 2001. A worldwide monograph of *Gentiana* [M]. Beijing: Science Press.
- HO TN, LIU SW, 2015. A worldwide monograph of *Swertia* and its allies [M]. Beijing: Science Press.
- HOSTETLER GL, RALSTON RA, SCHWARTZ SJ, 2017. Flavones: food sources, bioavailability, metabolism, and bioactivity [J]. *Adv Nutr*, 8(3): 423-435.
- JAAKOLA L, 2013. New insights into the regulation of anthocyanin biosynthesis in fruit [J]. *Trends Plant Sci*, 18(9): 477-483.
- KANG JY, CHOI HI, IM MY, et al., 2002. *Arabidopsis* basic leucine zipper proteins that mediate stress-responsive abscisic acid signaling [J]. *Plant Cell*, 14(2): 343-357.
- KANG YJ, YANG DC, KONG L, et al., 2017. CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features [J]. *Nucl Acid Res*, 45(W1): W12-W16.
- KOONIN EV, FEDOROVA ND, JACKSON JD, et al., 2004. A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes [J]. *Genome Biol*, 5(2): R7.
- LI AM, ZHANG JY, ZHOU ZY, 2014. PLEK: a tool for

- predicting long non-coding RNAs and messenger RNAs based on an improved *k*-mer scheme [J]. BMC Bioinf, 15: 311.
- LI J, HARATA-LEE Y, DENTON MD, et al., 2017. Long read reference genome-free reconstruction of a full-length transcriptome from *Astragalus membranaceus* reveals transcript variants involved in bioactive compound biosynthesis [J]. Cell Discov, 3: 17031.
- LI YL, ZHANG HM, CUI YG, et al., 2020. Analysis on SSR information in full-length transcriptome and development of molecular markers in *Allium tuberosum* [J]. Acta Hort Sin, 47(4): 759–768. [李延龙, 张华敏, 崔蕴刚, 等, 2020. 韭菜全长转录组 SSR 信息分析及分子标记开发 [J]. 园艺学报, 47(4): 759–768.]
- LIU JQ, HE TN, 1996. The embryological studies of *Comastoma pulmonarium* (Gentianaceae) [J]. Acta Phytotaxon Sin, 34(6): 577–585. [刘健全, 何廷农, 1996. 喉毛花的胚胎学研究 [J]. 植物分类学报, 34(6): 577–585.]
- LIU Q, ZHANG GY, CHEN SY, 2000. The structure and regulatory role of plant transcription factors [J]. Chin Sci Bull, 45(14): 1465–1474. [刘强, 张贵友, 陈受宜, 2000. 植物转录因子的结构与调控作用 [J]. 科学通报, 45(14): 1465–1474.]
- LIU Z, QI YY, WU HY, et al., 2021. Chemical constituents of Tibetan medicine *Comastoma pedunculatum* whole herb [J]. J Chin Med Mat, 44(6): 1397–1404. [刘真, 祁艳艳, 吴海燕, 等, 2021. 藏药长梗喉毛花的活性成分研究 [J]. 中药材, 44(6): 1397–1404.]
- LIU XC, ZHAO Y, ZENG QY, et al., 2016. Determination of five effective components in *Comastoma pulmonarium* by HPLC [J]. Chin J Trad Chin Med Pharm, 31(9): 3760–3762. [刘小翠, 赵英, 曾擎屹, 等, 2016. HPLC 法测定喉毛花植物中 5 种有效成分的含量 [J]. 中华中医药杂志, 31(9): 3760–3762.]
- MEI Y, LI XR, CAI SK, et al., 2021. Full-length transcriptome analysis of a homology of medicine and food of *Pueraria thomsonii* [J]. Acta Agric Boreal-Sin, 36(5): 10–17. [梅瑜, 李向荣, 蔡时可, 等, 2021. 药食同源植物甘葛藤的全长转录组分析 [J]. 华北农学报, 36(5): 10–17.]
- MINORU K, SUSUMU G, SHUICHI K, et al., 2004. The KEGG resource for deciphering the genome [J]. Nucl Acid Res, 32(Database issue): D277–D280.
- QIAO YQ, CUI BS, TANG L, et al., 2012. Chemical constituents of *n*-BuOH extract of *Comastoma pedunculatum* [J]. Chin J Chin Mat Med, 37(16): 2360–2365. [乔涌起, 崔保松, 唐丽, 等, 2012. 长梗喉毛花正丁醇部位化学成分研究 [J]. 中国中药杂志, 37(16): 2360–2365.]
- SALMELA L, RIVALS E, 2014. LoRDEC: Accurate and efficient long read error correction [J]. Bioinformatics, 30(24): 3506–3514.
- SUN L, LUO HT, BU DC, et al., 2013. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts [J]. Nucl Acid Res, 41(17): e166.
- TANG L, JIN ZN, MEN MJ, et al., 2007. Recent progress and utilization of tibetan medicine Zangyinchen [J]. J CUN (Nat Sci Ed), 2(16): 176–178. [唐丽, 金振南, 门美佳, 等, 2007. 藏药藏茵陈的研究进展及开发利用 [J]. 中央民族大学学报(自然科学版), 2(16): 176–178.]
- TATUSOV RL, FEDOROVA ND, JACKSON JD, et al., 2003. The COG database: an updated version includes eukaryotes [J]. BMC Bioinf, 4: 41.
- THOLL D, 2015. Biosynthesis and biological functions of terpenoids in plants [J]. Adv Biochem Eng Biotechnol, 148: 63–106.
- WAN HF, JIA XW, ZOU PF, et al., 2019. The single-molecule long-read sequencing of *Scylla paramamosain* [J]. Sci Rep, 9(1): 1–11.
- WANG B, TSENG E, REGULSKI M, et al., 2016. Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing [J]. Nat Commun, 7: 11708.
- WANG RX, LI C, 2019. Application of full-length transcriptome sequencing in non-model plant transcriptome research [J]. Mol Plant Breed, 17(2): 502–508. [王瑞娴, 李川, 2019. 全长转录组测序技术在非模式植物转录组学研究中的应用 [J]. 分子植物育种, 17(2): 502–508.]
- XIE DM, YU NJ, HUANG LQ, et al., 2017. Next generation sequencing and transcriptome analysis of root bark from *Paeonia suffruticosa* cv. Feng Dan [J]. Chin J Chin Mat Med, 42(15): 2954–2961. [谢冬梅, 俞年军, 黄璐琦, 等, 2017. 基于高通量测序的药用植物“凤丹”根皮的转录组分析 [J]. 中国中药杂志, 42(15): 2954–2961.]
- XU JH, ZENG Q, YE FY, et al., 2021. Genetic relationship of *Populus qiongdaoensis* in *Populus* based on full-length transcriptome sequences, nuclear genes and chloroplast genes [J]. J Beijing For Univ, 43(10): 28–37. [徐家洪, 曾晴, 叶富余, 等, 2021. 基于全长转录组序列、核基因与叶绿体基因分析琼岛杨在杨属的亲缘关系 [J]. 北京林业大学学报, 43(10): 28–37.]
- XU ZC, PETERS RJ, WEIRATHER J, et al., 2015. Full-length transcriptome sequences and splice variants obtained by a combination of sequencing platforms applied to different root tissues of *Salvia miltiorrhiza* and tanshinone biosynthesis [J]. Plant J, 82(6): 951–961.
- YAN X, LIU HQ, ZOU YQ, et al., 2008. Physiological activity and research advance in synthesis of flavonoids [J]. Chin J Org Chem, 28(9): 1534–1544. [延玺, 刘会青, 邹永青, 等, 2008. 黄酮类化合物生理活性及合成研究进展 [J]. 有机化学, 28(9): 1534–1544.]

- YANG QS, ZHAO Y, ZHANG HH, et al., 2013. Geographic distribution and utilization of the gentianaceae endemic Zangyinchen medicinal plant in northwest Yunnan [J]. J Yunnan Minzu Univ (Nat Sci Ed), 22(3): 176-181. [杨青松, 赵艳, 张海浩, 等, 2013. 龙胆科藏茵陈类民族药原植物在滇西北的地理分布与资源利用 [J]. 云南民族大学学报 (自然科学版), 22(3): 176-181.]
- YANG Y, GAO SQ, TANG YM, et al., 2009. Advance of bZIP transcription factors in plants [J]. J Trit Crop, 29(4): 730-737. [杨颖, 高世庆, 唐益苗, 等, 2009. 植物 bZIP 转录因子的研究进展 [J]. 麦类作物学报, 29(4): 730-737.]
- YU JY, XIA MZ, XU H, et al., 2022. Comparative transcriptome analysis of three *Artemisia* species in Qinghai Tibet Plateau [J]. Bull Bot Res, 42(2): 200-210. [余静雅, 夏铭泽, 徐浩, 等, 2022. 青藏高原地区 3 种蒿属植物转录组比较分析 [J]. 植物研究, 42(2): 200-210.]
- ZHAO L, GAO LP, WANG HX, et al., 2013. The R2R3-MYB, bHLH, WD40, and related transcription factors in flavonoid biosynthesis [J]. Funct Integr Genomics, 13(1): 75-98.
- ZHAO LY, CAO SY, LONG YS, et al., 2019. Applications and research progresses of full-length transcriptome sequencing in plants [J]. J Plant Genet Resour, 20(6): 1390-1398. [赵陆滢, 曹绍玉, 龙云树, 等, 2019. 全长转录组测序在植物中的应用研究进展 [J]. 植物遗传资源学报, 20(6): 1390-1398.]
- ZHAO XY, WANG Q, YANG S, et al., 2020. The transcriptome sequencing and analysis of the medicinal plant *Apocynum venetum* [J]. Mol Plant Breed, 18(8): 2503-2511. [赵雪艳, 王琪, 杨莎, 等, 2020. 药用植物罗布麻的转录组测序及分析 [J]. 分子植物育种, 18(8): 2503-2511.]
- ZHANG C, HU LJ, WANG Y, 2014. Reproductive allocation of three sympatric *Comastoma* (Gentianaceae) species [J]. Guihaia, 34(1): 38-45. [张婵, 胡莉娟, 王赞, 2014. 三种同域分布喉毛花的繁殖分配 [J]. 广西植物, 34(1): 38-45.]
- ZHANG JC, 2020. Identification and functional analysis of candidate genes foe seed shattering in *Elymus sibiricus* based on transcriptome sequencing [D]. Lanzhou: Lanzhou University: 1-2. [张俊超, 2020. 基于转录组测序挖掘老芒麦落粒候选基因及其功能分析 [D]. 兰州: 兰州大学: 1-2.]
- ZHANG N, SUN GL, DAI JG, et al., 2013. Sequencing and analysis of the transcriptome of *Ginkgo biloba* L. cells [J]. Chin Biotechnol, 33(5): 112-119. [张楠, 孙桂玲, 戴均贵, 等, 2013. 银杏细胞转录组高通量测序及分析 [J]. 中国生物工程杂志, 33(5): 112-119.]
- ZHANG QQ, ZHU JH, NI YM, et al., 2011. The structure and function of plant bHLH transcription factors [J]. J Trop Subtrop Bot, 19(1): 84-90. [张全琪, 朱家红, 倪燕妹, 等, 2011. 植物 bHLH 转录因子的结构特点及其生物学功能 [J]. 热带亚热带植物学报, 19(1): 84-90.]
- ZHANG Y, XIA MZ, ZHANG FQ, 2020. Transcriptome analysis for medicinal plant *Anisodus tanguticus* [J]. Bull Bot Res, 40(3): 458-467. [张雨, 夏铭泽, 张发起, 2020. 药用资源植物山莨菪的转录组信息分析 [J]. 植物研究, 40(3): 458-467.]
- ZHANG Y, YU JY, XIA MZ, et al., 2021. Plastome sequencing reveals phylogenetic relationships among *Comastoma* and related taxa (Gentianaceae) from the Qinghai-Tibetan Plateau [J]. Ecol Evol, 11(22): 16034-16046.
- ZHANG ZJ, LIU YR, ZHANG SJ, et al., 2020. The principle of the third generation sequencing technology and its application in biological field [J]. Chin J Anim Sci, 56(6): 11-15. [张子敬, 刘燕蓉, 张顺进, 等, 2020. 第三代测序技术的方法原理及其在生物领域的应用 [J]. 中国畜牧杂志, 56(6): 11-15.]
- ZHENG Y, JIAO C, SUN HH, et al., 2016. ITAK: A program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases [J]. Mol Plant, 9(12): 1667-1670.
- ZHONG GY, WANG CH, LIU X, et al., 2010. The resources and usage status of the commonly used Tibetan medicinal crop "Dida" [J]. World Sci Technol-Mod Trad Chin, 12(1): 122-128. [钟国跃, 王昌华, 刘翔, 等, 2010. 常用藏药“蒂达(藏茵陈)”的资源与使用现状调查 [J]. 世界科学技术: 中医药现代化, 12(1): 122-128.]
- ZHU FY, CHEN MX, YE NH, et al., 2017. Proteogenomic analysis reveals alternative splicing and translation as part of the abscisic acid response in *Arabidopsis* seedlings [J]. Plant J, 91(3): 518-533.
- ZHU XZ, XIA LF, CHEN LB, et al., 2018. Full-length transcriptome analysis of protected cultivation 'Yuncha 1' (*Camellia sinensis* var. *assamica*) [J]. J Tea Sci, 38(2): 193-201. [朱兴正, 夏丽飞, 陈林波, 等, 2018. 保护品种云茶 1 号茶树全长转录组测序分析 [J]. 茶叶科学, 38(2): 193-201.]

(责任编辑 邓斯丽)