

DOI: 10.11931/guihaia.gxzw201909063

夏铭泽, 张雨, 余静雅, 等. 多裂骆驼蓬叶片转录组分析 [J]. 广西植物, 2021, 41(4): 503-513.

XIA MZ, ZHANG Y, YU JY, et al. Transcriptome analysis for leaves of *Peganum multisectum* [J]. *Guihaia*, 2021, 41(4): 503-513.

## 多裂骆驼蓬叶片转录组分析

夏铭泽<sup>1,2</sup>, 张雨<sup>1,2</sup>, 余静雅<sup>1,2</sup>, 张发起<sup>1\*</sup>

(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 100049)

**摘要:** 多裂骆驼蓬为西北荒漠地区常见植物, 具有抗风固沙、防止水土流失、抑菌杀虫和抗肿瘤等功效。为了增加骆驼蓬属植物开发利用的强度, 弥补其基因功能、代谢通路等分子生物学研究层面的空缺, 该文利用 Illumina 高通量测序平台对多裂骆驼蓬叶片进行转录组测序, 根据测序结果进行转录组数据拼接、功能注释、序列水平和表达水平等分析。结果表明: 共获得 7 723 653 900 bp 核苷酸序列信息, 拼接组装得到 78 641 条 Unigene, 预测出 55 535 条 CDS 序列。与多个数据库对比后, 获得 33 184 个 NR 数据库注释信息; 31 835 个 GO 数据库注释信息; 17 206 个 KOG 数据库 Unigene 功能分类信息; 7 617 个 KEGG 数据库代谢通路注释信息。对单核苷酸多态性位点和微卫星信息进行检测分析, 共发现 86 113 个 SNP 位点, 6 987 个 SSR 信息。该研究首次获得并分析了多裂骆驼蓬的转录组数据, 通过基因比对、CDS 预测、通路注释、SNP 检测和 SSR 检测等方法, 对该植物的基因、通路以及分子标记等方面有了初步的认识, 为本种植物后续研究及资源的开发利用奠定了基础。

**关键词:** 多裂骆驼蓬, 转录组分析, 基因注释, SNP, SSR

中图分类号: Q943 文献标识码: A 文章编号: 1000-3142(2021)04-0503-11

## Transcriptome analysis for leaves of *Peganum multisectum*

XIA Mingze<sup>1,2</sup>, ZHANG Yu<sup>1,2</sup>, YU Jingya<sup>1,2</sup>, ZHANG Faqi<sup>1\*</sup>

(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 100049, China)

**Abstract:** *Peganum multisectum* is a common plant in arid-desert areas of Northwest China, which has the effects of resisting wind and sand fixation, preventing soil erosion, inhibiting bacteria and insects, and resisting tumor. But the lack of molecular biology research on this plant, such as gene function and metabolic pathway, leads to its underutilization. Here in this study, the transcriptomes of *P. multisectum* leaves were sequenced by the platform of Illumina. According to the sequencing results, we analyzed splicing, functional annotation, sequence level and expression level of the transcriptome data. Then we obtained a total of 7 723 653 900 bp nucleotide sequence information, assembled 78 641 Unigene sequences and predicted 55 535 CDS sequences. After comparing with several databases, 33 184 NR database annotation information, 31 835 GO database annotation information, 17 206 KOG database Unigene functional classification information and 7 617 KEGG database metabolic pathway annotation

收稿日期: 2020-02-06

基金项目: 国家自然科学基金(31110103911); 青海省应用基础研究计划项目(2019-ZJ-7042) [Supported by the National Natural Science Foundation of China (31110103911); Basic Research Program of Qinghai Province (2019-ZJ-7042)].

作者简介: 夏铭泽(1994-), 博士研究生, 主要从事植物系统与进化研究, (E-mail) xiamingze16@mails.ucas.ac.cn.

\*通信作者: 张发起, 博士, 副研究员, 主要从事植物多样性研究, (E-mail) fqzhang@nwipb.cas.cn.

information were obtained. In addition, we detected a total of 86 113 single nucleotide polymorphism sites and 6 987 microsatellite information. In this study, transcriptome data of *P. multisectum* were obtained and analyzed for the first time. These analyses not only shed light on the molecular information of *P. multisectum*, but also provide a certain reference for the post research, development and utilization of this plant.

**Key words:** *Peganum multisectum*, transcriptome analysis, gene annotation, SNP, SSR

1753 年植物学家林奈 (Carl Linnaeus) 以骆驼蓬 (*Peganum harmala*) 为模式种建立了骆驼蓬属 (*Peganum* L.) (徐朗然和黄成就, 1993)。该属在《中国植物志》中属于蒺藜科 (Zygophyllaceae) (徐朗然和黄成就, 1993), 在 *Flora of China* 中属于骆驼蓬科 (Peganaceae), 在 APG IV 系统中属于白刺科 (Nitrariaceae) 骆驼蓬亚科 (Subfam. Peganoideae Engl.) (Angiosperm Phylogeny Group et al., 2016)。骆驼蓬属植物为多年生草本, 全世界共 6 种, 中国有 3 种, 即骆驼蒿 (*Peganum nigellastrum*)、骆驼蓬 (*P. harmala*) 和多裂骆驼蓬 (*P. multisectum*), 主要分布于新疆、宁夏、青海、甘肃和内蒙古的荒漠或半荒漠地区 (徐朗然和黄成就, 1993)。多裂骆驼蓬是中国特有种 (段金廛等, 1998), 该植物对各种生境的适应力强于骆驼蒿和骆驼蓬 (刘媵心, 1995), 具有旱生植物的典型特征, 如根系发达, 叶片深裂且裂片较细, 有较强的耐旱特性。此外, 多裂骆驼蓬植物体生长繁茂, 在改善西北荒漠环境中发挥着抗风固沙和防止水土流失的重要作用 (马骥和王勋陵, 1998; 程琳等, 2018)。该种资源丰富, 仅甘肃省年产量超过 1 亿 Kg (樊峥嵘和姚新生, 1992)。近年来的研究表明, 骆驼蓬属植物具有抑菌杀虫 (薛林贵等, 2005)、抗肿瘤 (陈豫等, 2015) 等多种药理学活性; 提取物和生物碱对多种植物种子的萌发有抑制作用 (刘建新, 2003; 刘建新等, 2008)。段金廛等 (1998) 从多裂骆驼蓬中分离并鉴定了 22 种化合物, 刘建新等 (2011) 研究显示外源一氧化氮可以保护盐胁迫下多裂骆驼蓬幼苗的光合系统。目前, 对多裂骆驼蓬的研究主要集中在药理活性、种子萌发及幼苗生长方面。已报道的分子序列仅见于 Zhao 等 (2011) 利用 *trnL-F* 片段和 *psbA-trnH* 片段对骆驼蓬属植物的鉴定, 基因注释信息亦不完善, 阻滞了与药物活性成分合成相关的代谢通路的研究。因此, 亟需利用基因组或转录组数据来拓展分子生物学信息。

转录组 (transcriptome) 是指生物体的细胞或组织在特定时空条件下转录出的 RNA 的总和 (张春

兰等, 2012)。较基因组而言, 转录组仅涉及被转录的基因, 可以更有效地针对功能基因及其代谢通路进行研究 (周华等, 2012)。随着近年来高通量测序技术的发展, 其测序时间和测序成本大大降低 (Shendure & Ji, 2008), 越来越多的学者将高通量测序技术应用到转录组研究中, 从而获得来自不同基因的海量 RNA 序列数据。如叶兴状等 (2019) 对濒危珍稀植物半枫荷 (*Semiliquidambar cathayensis*) 进行转录组分析, 发现了控制药效合成的转录因子家族; Zhang et al. (2014) 对青藏高原特有植物唐古红景天 (*Rhodiola tangutica*) 植物进行比较转录组分析, 为解释该植物可用于预防高原反应的机制提供了分子水平的依据; 李彦等 (2018) 分析了山地虎耳草 (*Saxifraga sinomontana*) 的转录组微卫星位点信息, 为该物种遗传多样性研究提供了理论依据; 付苏宏等 (2019) 研究了菊叶香藜 (*Dysphania schraderilana*) 转录组数据库中的 *FPPS* 基因, 对该植物倍半萜类化合物的生物合成奠定了基础。目前, 已有部分具药用价值的植物完成转录组测序工作, 如人参 (*Panax ginseny*) (邹丽秋等, 2016)、三七 (*P. notoginseng*) (黄勋等, 2017)、甘草 (*Glycyrrhiza uralensis*) (张春荣等, 2015)、乌头 (*Aconitum carmichaelii*) (张大燕等, 2017) 和金铁锁 (*Psammosilene tunicoides*) (孟文俊, 2019) 等, 为进一步研究药用植物的功能基因及代谢通路提供了分子基础。

骆驼蓬属植物资源丰富, 为增加本属植物开发利用的强度, 应对其基因功能、代谢通路等进行深入研究。因此, 本文选取多裂骆驼蓬叶片为研究材料, 利用 Illumina Hiseq 2500 高通量测序平台对其进行首次转录组测序, 建立该物种转录组数据库, 通过生物信息学方法对 Unigene 进行功能注释、代谢途径及 SSR 检测等分析, 挖掘多裂骆驼蓬具有药效和抗干旱作用的功能基因, 为进一步研究该植物代谢通路和遗传多样性奠定基础, 也为今后本属植物的开发利用和保护提供理论依据。

# 1 材料与amp;方法

## 1.1 材料

多裂骆驼蓬植株采自青海省共和县铁盖乡(36°06'56.61" N, 100°41'41.74" E, 2 594 m)。将采集的成熟植物叶片放入液氮中快速冷冻,返回实验室后放入-80 °C超低温冰箱保存。凭证标本(Zhang2018013)存放在中国科学院西北高原生物研究所青藏高原生物标本馆(HNWP)中。

## 1.2 方法

**1.2.1 RNA 提取** 从多裂骆驼蓬的叶片材料中提取总 RNA,用琼脂糖凝胶电泳分析 RNA 的降解程度以及是否有污染;用 Nanodrop 2000 初步检测 RNA 的纯度(A260/A280 应在 1.9~2.0 范围内);使用 Agilent 2100 测定样品的 RIN 值以确定 RNA 完整性。

**1.2.2 拼接组装** 利用 Illumina Hiseq 2500 高通量测序平台进行测序,得到的原始图像数据由 CASAVA 碱基识别(base calling)后转化为原始测序数据(Raw reads)。该数据文件包含测序得到的碱基序列(reads)信息及其对应的碱基测序质量信息。质量评估后对所得的原始测序数据进行过滤:去除接头(adapter)序列;去除含有未知碱基的序列;从序列两端起始,去除低质量的碱基。过滤后的测序数据为 Clean reads,使用软件 Trinity (Grabherr et al., 2011) (版本为 2.4.0;参数设置为 min\_kmer\_cov = 2,其余为默认设置)将该数据组装成转录本,并取每条基因中最长的转录本作为 Unigene,以此作为后续分析的参考序列。

**1.2.3 功能注释** 使用 NCBI BLAST+ (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) 将 Unigene 序列与 CDD、KOG、COG、NR、NT、PFAM、Swissprot 和 TrEMBL 数据库比对,获得基因的功能注释信息。首先根据 Swissprot 数据库和 TrEMBL 数据库的蛋白注释结果,结合 Uniprot (UniProt, 2018) 的注释信息得到 GO 注释。使用软件 KAAS (KEGG Automatic Annotation Server) (Moriya, 2007) 将 Unigene 序列与 KEGG 基因数据库进行 BLAST 或 GHOST 比较,以此来获得基因的功能注释信息。首先根据 NR、Swissprot、TrEMBL 的最佳比对结果按其优先级顺序确定 Unigene 的 ORF 读码框;然后根据标准密码子表确定其 CDS 序列(Coding Sequence, 氨基酸序列),同时将未比对上的

Unigene 通过 TransDecoder 软件预测其 CDS 序列。

**1.2.4 SNP 和 SSR 的筛选与统计** 以组装好的 Unigene 作为参考序列,使用 BCFtools (版本为 2.4.3) (Quinlan, 2010) 找出其中的 SNP 位点,参数为质量值大于 20 且覆盖度大于 8,对筛选出的 SNP 突变类型进行统计分析。使用 MISA (版本为 1.0.1) (Thiel, 2003) 检测 Unigene 序列中的 SSR 信息,检测的类型包括完美型(perfect)及复合型(compound)的 SSR。微卫星各重复单元的筛选标准为二核苷酸 SSR 至少重复 6 次,三核苷酸 SSR 至少重复 5 次,四核苷酸至少重复 5 次,五核苷酸至少重复 5 次,六核苷酸至少重复 5 次。使用 Excle (版本为 Microsoft Office 2016) 软件对 SSR 的类型、数量等进行统计分析。

# 2 结果与分析

## 2.1 转录组数据组装及 Unigene 获取

对多裂骆驼蓬叶片 mRNA 进行测序,共获得 51 491 026 条 Raw reads,包含 7 723 653 900 bp 的核苷酸序列信息,平均每条 read 的长度为 150 bp,长度大于 30 bp 序列的占比为 93.76%,GC 百分比为 47.65%。对 Raw reads 进行过滤,获得 50 104 364 条 Clean reads,包含 7 257 449 296 bp 的核苷酸序列信息,平均每条 read 的长度为 144.85 bp,长度大于 30 bp 序列的占比为 95.06%,GC 百分比为 47.46%。

用 Trinity 软件对 Clean reads 进行 *de novo* 组装后获得 148 317 个 Transcript,去冗余后获得 78 641 条 Unigene,其总的核苷酸数分别为 117 421 439、52 289 971 个(表 1)。Unigene 长度在 200~300 nt 之间的有 31 316 条,占比 39.82%;在 300~2 000 nt 之间的有 42 524 条,占比 54.07%;长度超过 2 000 nt 的有 4 801 条,占比 6.10%(图 1:A)。同时,对多裂骆驼蓬 Unigene 的编码序列进行预测,共获得 55 535 条 CDS 序列。CDS 序列长度在 100~200 nt 之间的有 12 541 条,占比 22.58%;在 200~300 nt 之间的有 15 311 条,占比 27.57%;在 300~2 000 nt 之间的有 26 416 条,占比 47.57%;长度超过 2 000 nt 的有 1 267 条,占比 2.28%(图 1:B)。

## 2.2 转录组基因功能注释

将 Unigene 分别与 CDD、KOG、NR、NT、PFAM、Swissprot、TrEMBL、GO 和 KEGG 数据库进行比对

表 1 多裂骆驼蓬转录组数据拼接结果

Table 1 Assembly result of transcriptome data from *Peganum multisectum*

数据类型 Data type	数量 Number	≥ 500 nt	≥ 1 000 nt	N <sub>50</sub>	N <sub>90</sub>	最大长度 Max length (bp)	最小长度 Min length (bp)	总长 Total length (bp)	平均长度 Average length (bp)
Transcript	148 317	74 914	41 673	1 259	319	8 787	201	117 421 439	791.69
Unigene	78 641	29 091	16 071	1 154	260	8 787	201	52 289 971	664.92

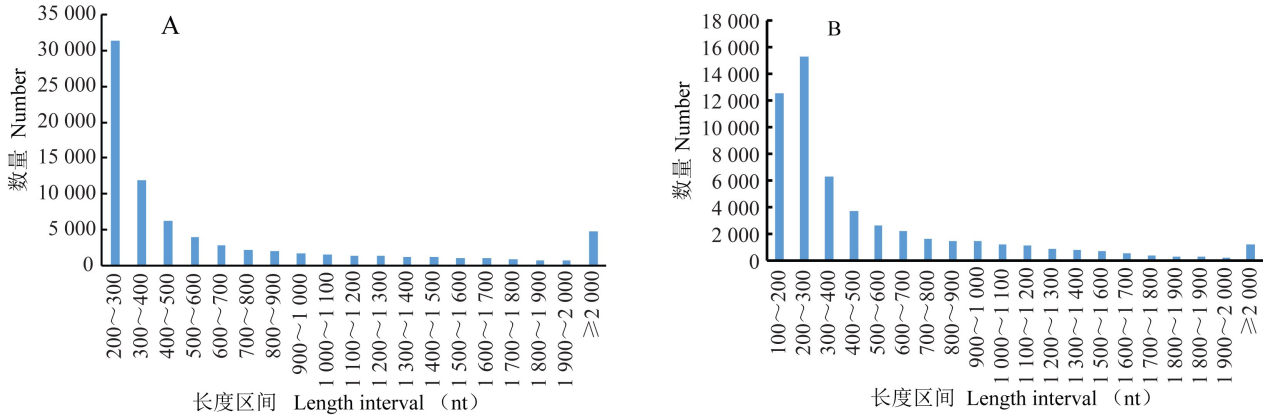


图 1 多裂骆驼蓬 Unigene 分布图(A) 与 CDS 长度分布图(B)

Fig. 1 Length distribution of Unigene (A) and CDS of *Peganum multisectum* (B)

并注释基因功能(表 2)。其中,注释到 NR 数据库、TrEMBL 数据库和 GO 数据库的基因数目最多,分别有 33 184、32 887、31 835 个,占总基因数目的 42.2%、41.82%、40.48%;注释到 KEGG 数据库的基因数目最少,为 3 875 个,仅占总基因数目的 2.97%。总体而言,共有 38 598 个 Unigene 基因经过多数据库比对后成功注释,仍有 40 043 个 Unigene 基因未能获取注释信息。

### 2.3 NR 数据库 Unigene 近似物种比对

将 Unigene 与 NR 数据库进行比对,结果显示有 33 184 个 Unigene 获得注释信息,占 Unigene 总数的 42.2%。通过与 NR 库的比对,揭示了多裂骆驼蓬转录组序列与库中物种的近似情况,结果显示该物种与甜橙 (*Citrus sinensis*) 的相似序列最多,有 7 078 条,占 NR 库注释的 Unigene 总数的 21.33%;其次为克莱门柚 (*C. clementina*),有 4 827 条,占 NR 库注释的 Unigene 总数的 14.55%。此外,部分物种还与多裂骆驼蓬有大量基因序列相类似,如可可 (*Theobroma cacao*)、葡萄 (*Vitis vinifera*)、麻疯树 (*Jatropha curcas*)、蓖麻 (*Ricinus communis*)、枣 (*Ziziphus jujuba*)、毛果杨 (*Populus*

表 2 Unigene 注释结果统计

Table 2 Statistics of Unigene annotation results

数据库 Database	基因数量 Number of Unigenes	百分比 Percentage (%)
注释到 CDD Annotated in CDD	22 911	29.13
注释到 KOG Annotated in KOG	17 206	21.88
注释到 NR Annotated in NR	33 184	42.2
注释到 NT Annotated in NT	20 527	26.1
注释到 PFAM Annotated in PFAM	17 042	21.67
注释到 Swissprot Annotated in Swissprot	28 796	36.62
注释到 TrEMBL Annotated in TrEMBL	32 887	41.82
注释到 GO Annotated in GO	31 835	40.48
注释到 KEGG Annotated in KEGG	3 875	4.93
至少注释到一个数据库 Annotated in at least one databases	38 598	49.08
注释到所有数据库 Annotated in all database	2 338	2.97
Unigene 总数 Total Unigenes	78 641	100



*trichocarpa*)、雷蒙德氏棉(*Gossypium raimondii*)和大麦(*Hordeum vulgare*),这些物种与多裂骆驼蓬相似的基因序列共有 7 218 条,共占 NR 库注释的 Unigene 总数的 21.75%。其余 42.37% 的注释 Unigene 分布于其他 561 个物种中(图 2)。

## 2.4 Gene Ontology 数据库 Unigene 功能分析

在 GO 数据库中比对骆驼蓬转录组 Unigene,有 31 835 条 Unigene 获得 210 748 条注释信息。所有注释信息分为三大类,即生物学过程(biological process)、细胞组分(cellular component)和分子功能(molecular function)。所有大类可细分为 67 个二级分类。对 Unigene 在二级分类中的分布情况进行统计分析,结果显示细胞(cell)和细胞部分(cell part)类型的 Unigene 最多,分别为 23 443、23 399 个,分别占 GO 注释信息总数的 73.64%、73.50%。此外,细胞过程(cellular process)、连接(binding)、细胞器(organelle)、代谢过程(metabolic process)和细胞膜(membrane)类型的 Unigene 也较多,分别有 21 154 (66.45%)、18 776 (58.98%)、18 689 (58.71%)、18 477 (58.04%)和 10 999 (34.55%)条注释信息。化学引诱剂活性(chemoattractant activity)、生物相(biological phase)和形态发生素(morphogen activity)类型的注释信息较少(图 3)。

## 2.5 KOG 数据库 Unigene 功能分类

在 KOG 数据库中比对骆驼蓬转录组 Unigene,结果显示共 17 206 条 Unigene 获得注释信息,被分为 25 类(图 4)。其中,一般功能预测基因(general function prediction only)、翻译后修饰、蛋白质折叠和分子伴侣(posttranslational modification, protein turnover, chaperones)、信号传导机制(signal transduction mechanisms)获得注释最多,分别有 2 179、2 049 和 1 871 条。细胞活性(cell motility)获得注释最少,仅 6 条(图 4)。

## 2.6 KEGG 数据库 Unigene 代谢通路分析

使用 KEGG 数据库对 Unigene 序列进行比对注释,代谢通路分析结果显示多裂骆驼蓬转录组中共有 3 875 个 Unigene 获得 7 617 条代谢通路注释信息。这些通路信息可分为四大类,分别为细胞进程(cellular processes)、环境信息处理(environmental information processing)、遗传信息处理(genetic information processing)和新陈代谢(metabolism)。这四大类可进一步分为 23 个小类

(图 5)。其中,翻译(translation)和信号传导(signal transduction)获得注释信息最多,分别有 662 条和 590 条;信号分子和相互作用(signaling molecules and interaction)及膜运输(membrane transport)的注释最少,有 2 条和 21 条。

所有注释的代谢通路信息中,共有 240 个 Unigene 与药物活性成分合成相关。其中,51 个 Unigene 参与了萜类骨架生物合成通路(KEGG 数据库通路 ID:ko00900),5 个 Unigene 参与了单萜类生物合成通路(KEGG 数据库通路 ID:ko00902),9 个 Unigene 参与了二萜类生物合成通路(KEGG 数据库通路 ID:ko00904),9 个 Unigene 参与了倍半萜类和三萜类生物合成通路(KEGG 数据库通路 ID:ko00909),这些通路均与萜类物质合成相关。有 62 个 Unigene 参与了苯丙烷类生物合成通路(KEGG 数据库通路 ID:ko00940),这与苯丙素类物质合成相关。此外,还有 13 个 Unigene 参与了黄酮类生物合成通路(KEGG 数据库通路 ID:ko00941),16 个 Unigene 参与了易啉生物碱生物合成通路(KEGG 数据库通路 ID:ko00950),31 个 Unigene 参与了泛醌和其他萜类醌生物合成通路(KEGG 数据库通路 ID:ko00130),19 个 Unigene 参与了花生四烯酸类生物合成通路(KEGG 数据库通路 ID:ko00590)。

多裂骆驼蓬在干旱生境下仍有较强的适应力(刘媵心,1995)。通过转录组数据分析发现,多裂骆驼蓬叶片转录信息包含有多个脱落酸(abscisic acid, ABA)合成相关基因,即 *ZEP* (zeaxanthin epoxidase) 基因、*AAO3* (abscisic-aldehyde oxidase) 基因和 *PYL* (abscisic acid receptor) 基因。此外,还有茉莉酸(jasmonic acid, JA)合成相关基因,即 *LOX* (lipoxygenase) 基因、*TGA* (transcription factor) 基因、*JAR1* (jasmonic acid-amino synthetase) 基因、*AOS* (hydroperoxide dehydratase) 基因和 *MYC2* (transcription factor) 基因。已有研究表明,脱落酸和茉莉酸的合成可以增加植物的抗逆能力(张春荣等,2015)。

## 2.7 多裂骆驼蓬 Unigene 序列单核苷酸多态性检测分析

对多裂骆驼蓬 Unigene 序列进行检测,共发现 86 113 个单核苷酸多态性(SNP, single nucleotide polymorphisms)位点。对这些 SNP 位点进行类型统计,结果显示转换突变类型的 SNP 有 58 486 个

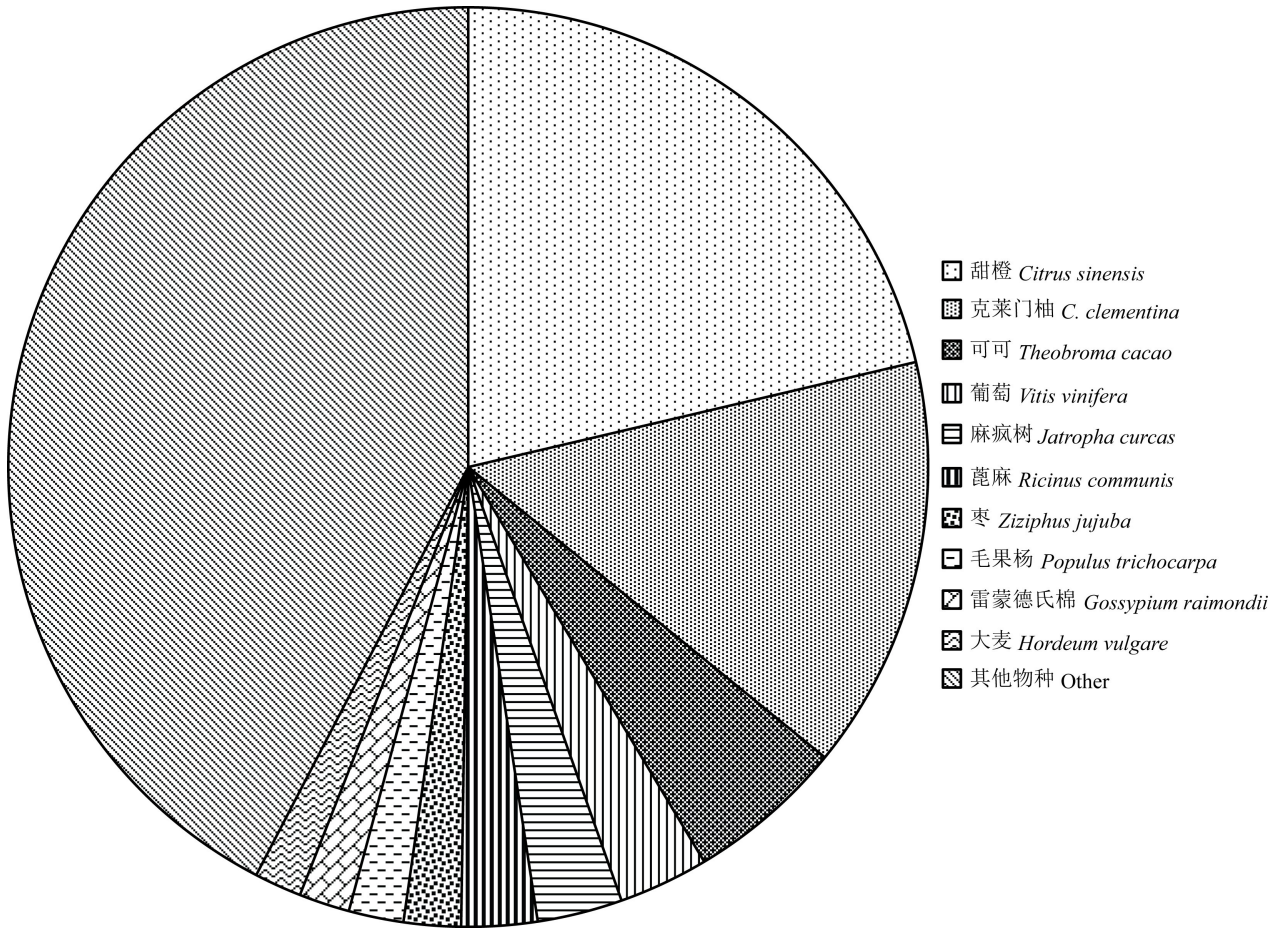


图 2 多裂骆驼蓬 NR 库注释出的物种分布图

Fig. 2 The distribution of NR annotated species of *Peganum multisectum*

(占 67.92%), 颠换突变类型的 SNP 有 27 627 个 (占 32.08%)。在转换突变类型中, 由鸟嘌呤转换为腺嘌呤的突变最多 (15 042 个), 其次为胞嘧啶转换为胸腺嘧啶的突变 (15 000 个), 表明 G→A 和 C→T 二者发生频率相差不大。在颠换突变类型中, 由腺嘌呤颠换为胸腺嘧啶的突变最多 (4 136 个), 而鸟嘌呤颠换为胞嘧啶的突变最少 (2 700 个) (图 6)。

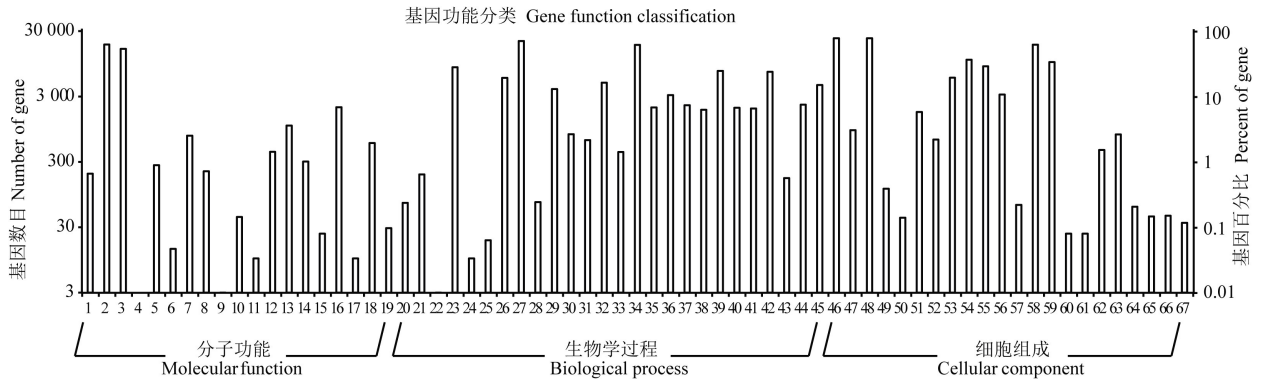
### 2.8 多裂骆驼蓬 Unigene 序列微卫星信息分析

利用软件 MISA 对多裂骆驼蓬 Unigene 进行微卫星位点 (SSR, simple sequence repeat) 进行检测, 共检测出 3 399 个 SSR。对 SSR 类型进行统计, 结果显示复合型 (compound) SSR 和完美型 (perfect) SSR 分别有 442 个和 2 957 个, 其发生频率为 4.32% (检测出的 SSR 数量与总序列数目的比值)。从分布情况来看, 多裂骆驼蓬转录组序列中

平均每 15.38 kb (序列总长度与 SSR 总数目的比值) 出现一个 SSR, 表明该物种转录组 SSR 数量较为丰富。在完全型 SSR 中, 三核苷酸重复占 45.01%; 二核苷酸重复次之, 占 38.22%; 四核苷酸重复、五核苷酸重复和六核苷酸重复分别占 2.53%、0.65% 和 0.59%。

### 3 讨论与结论

多裂骆驼蓬是西北荒漠地区的常见植物, 因其根系发达, 具有良好的水土保持作用 (马骥和王勋陵, 1998), 且该植物作为维族、蒙古族和藏族的常用草药 (徐小平等, 2008; 李凯等, 2015) 已有悠久的历史。多裂骆驼蓬在西北地区资源量巨大, 为了合理有效地利用这一植物资源, 我们通过转录组测序的手段以获取其序列信息及基因表达



1. 抗氧化活性; 2. 结合活性; 3. 催化活性; 4. 化学吸引活性; 5. 电子转移活性; 6. 金属伴侣活性; 7. 分子功能调节剂; 8. 分子转导活性; 9. 形态发生素活性; 10. 营养贮藏活性; 11. 蛋白质标记; 12. 信号转导活性; 13. 结构分子活性; 14. 转录因子活性与蛋白质结合; 15. 翻译调节活性; 16. 转运蛋白活性; 17. 通道调节活性; 18. 酶调节活性; 19. 受体调节活性; 20. 行为; 21. 生物黏附; 22. 生物相; 23. 生物调节; 24. 细胞聚集; 25. 细胞杀伤; 26. 细胞成分组织或生物合成; 27. 细胞过程; 28. 解毒作用; 29. 发育过程; 30. 生长; 31. 免疫系统过程; 32. 定位; 33. 转运; 34. 代谢过程; 35. 多组织进程; 36. 多细胞生物过程; 37. 生物过程负调控; 38. 生物过程正调控; 39. 生物过程的调节; 40. 繁殖; 41. 生殖过程; 42. 对刺激的反应; 43. 节律过程; 44. 信号传输; 45. 建立本地定位; 46. 细胞; 47. 细胞连接; 48. 细胞部分; 49. 细胞外基质; 50. 胞外基质成分; 51. 胞外区域; 52. 胞外区域部分; 53. 含蛋白复合体; 54. 膜; 55. 膜部分; 56. 膜封闭腔; 57. 拟核; 58. 细胞器; 59. 细胞器部分; 60. 其他有机体; 61. 其他有机体部分; 62. 超分子纤维; 63. 共质体; 64. 突触; 65. 突触部分; 66. 病毒; 67. 病毒部分。

1. Antioxidant activity; 2. Binding activity; 3. Catalytic activity; 4. Chemoattractant activity; 5. Electron transfer activity; 6. Metallochaperone activity; 7. Molecular function regulator; 8. Molecular transducer activity; 9. Morphogen activity; 10. Nutrient reservoir activity; 11. Protein tag; 12. Signal transducer activity; 13. Structural molecule activity; 14. Transcription factor activity, protein binding; 15. Translation regulator activity; 16. Transporter activity; 17. Channel regulator activity; 18. Enzyme regulator activity; 19. Receptor regulator activity; 20. Behavior; 21. Biological adhesion; 22. Biological phase; 23. Biological regulation; 24. Cell aggregation; 25. Cell killing; 26. Cellular component organization or biogenesis; 27. Cellular process; 28. Detoxification; 29. Developmental process; 30. Growth; 31. Immune system process; 32. Localization; 33. Locomotion; 34. Metabolic process; 35. Multi-organism process; 36. Multicellular organismal process; 37. Negative regulation of biological process; 38. Positive regulation of biological process; 39. Regulation of biological process; 40. Reproduction; 41. Reproductive process; 42. Response to stimulus; 43. Rhythmic process; 44. Signaling; 45. Establishment of localization; 46. Cell; 47. Cell junction; 48. Cell part; 49. Extracellular matrix; 50. Extracellular matrix component; 51. Extracellular region; 52. Extracellular region part; 53. Protein-containing complex; 54. Membrane; 55. Membrane part; 56. Membrane-enclosed lumen; 57. Nucleoid; 58. Organelle; 59. Organelle part; 60. Other organism; 61. Other organism part; 62. Supramolecular fiber; 63. Symplast; 64. Synapse; 65. Synapse part; 66. Virion; 67. Virion part.

图 3 多裂骆驼蓬 GO 功能统计图

Fig. 3 Statistical diagram of GO function of *Peganum multisectum*

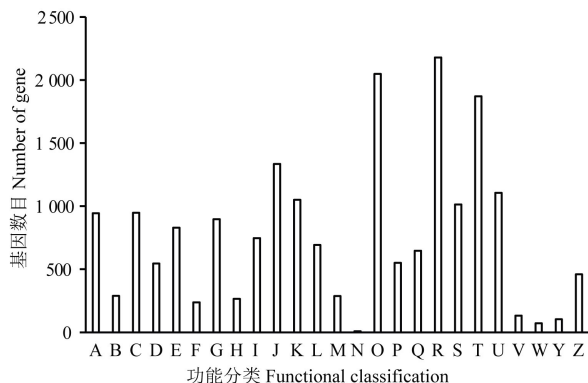
特征。

本研究通过对多裂骆驼蓬叶片进行转录组测序,共获得了 51 491 026 条 Raw reads 和 50 104 364 条 Clean reads。拼接组装得到 78 641 条 Unigene,  $N_{50}$  长度为 1 154, 平均长度为 664.92 bp, 比半枫荷 (叶兴状等, 2019)、胡卢巴 (*Trigonella foenum-graecum*) (Patel et al., 2014)、川芎 (*Ligusticum chuansiong*) (袁灿等, 2017)、狼毒 (*Stellera chamaejasme*) (杨艳芳等, 2017) 和东北红豆杉 (*Taxus cuspidata*) (吴琼等, 2012) Unigene 的平均长度更长, 说明本研究拼接所得长序列 Unigene 较多, 并对 Unigene 的编码序列进行预测, 共获得 55 535 条 CDS 序列。所得长序列 Unigene 的增多可能与测序物种的差异有关。

将 Unigene 序列与 CDD、KOG、COG、NR、NT、

PFAM、Swissprot、TrEMBL、GO 和 KEGG 共 10 个数据库比对, 共有 38 598 条 (49.08%) 序列经过多数数据库比对后成功注释, 但仍有 40 043 条 (50.92%) 未得到准确定位, 这一现象在许多物种的转录组结果中均有出现, 如樟树 (*Cinnamomum camphora*) (江香梅等, 2014)、黄秋葵 (*Abelmoschus esculentus*) (Schafleitner et al., 2013) 和紫背天葵 (*Begonia fimbristipula*) (张少平等, 2016)。这可能与某些 Unigene 片段长度太短、相关数据库基因注释信息不完善或者该物种存在新基因等因素有关。通过与 NR 库的比对, 显示多裂骆驼蓬与甜橙、克莱门柚、可可和葡萄等物种具有大量相似基因序列。多裂骆驼蓬与四个同科植物仅有 14 条相似基因序列, 分别为白刺 (*Nitraria tangutorum*) 7 条、*N. retusa* 4 条、骆驼蓬 2 条和小果白刺 (*N. sibirica*)





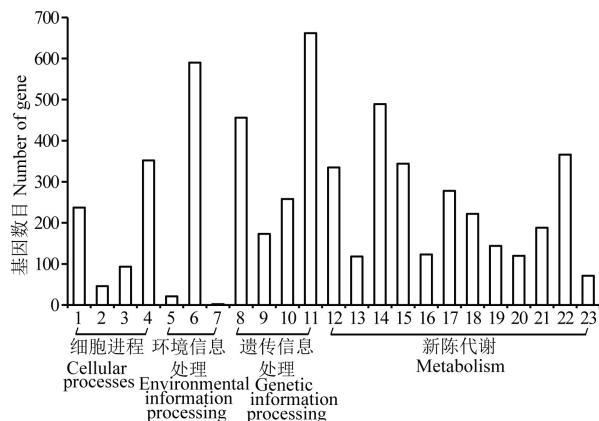
A. RNA 加工和修饰; B. 染色体结构和动力学; C. 能源生产与转化; D. 细胞周期调控, 细胞分裂, 染色体分离; E. 氨基酸转运和代谢; F. 核酸转运和代谢; G. 碳水化合物转运和代谢; H. 辅酶转运和代谢; I. 脂类转运和代谢; J. 翻译, 核糖体结构和生物发生; K. 转录; L. 复制, 重组和修饰; M. 细胞壁/细胞膜/细胞被膜生物发生; N. 细胞活性; O. 翻译后修饰, 蛋白质更新, 蛋白质伴侣; P. 无机离子转运和代谢; Q. 次生代谢物合成, 转运和代谢; R. 只有一般功能预测; S. 未知功能; T. 信号传递机制; U. 细胞间运输, 分泌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

图 4 多裂骆驼蓬 KOG 功能统计图

Fig. 4 Statistical diagram of KOG function of *Peganum multisectum*

1 条, 或与白刺科植物基因组、转录组数据严重缺乏有关。通过比对 GO 数据库, 获得 31 835 条 Unigene 的 210 748 个功能分类信息, 通过比对 KOG 数据库, 获得 17 206 个 Unigene 注释信息, 因此对多裂骆驼蓬 Unigene 的功能分布状况有了初步的了解。此外, 对拼接组装的 78 641 条 Unigene 进行代谢通路分析, 通过比对 KEGG 数据库最终获得 3 875 条 (4.93%) Unigene 的 7 617 条代谢通路注释信息, 注释类别可分为四大类, 其中定位到新陈代谢相关通路的基因数最多, 占总注释量的 47.15%, 表明多裂骆驼蓬代谢活动能力较强, 之后



1. 细胞生长和死亡; 2. 细胞运动; 3. 细胞群体; 4. 运输和分解代谢; 5. 膜运输; 6. 信号转导; 7. 信号分子和相互作用; 8. 依序折叠与降解; 9. 复制和修复; 10. 转录; 11. 翻译; 12. 氨基酸代谢; 13. 其他次级代谢产物合成; 14. 碳水化合物代谢; 15. 能量代谢; 16. 多糖合成与代谢; 17. 类脂(化合物)代谢作用; 18. 辅助因子和维生素的代谢; 19. 其他氨基酸的代谢; 20. 萜类化合物和聚酮化合物的代谢; 21. 核苷酸代谢; 22. 概观; 23. 异生素生物降解和代谢。

1. Cell growth and death; 2. Cell motility; 3. Cellular community; 4. Transport and catabolism; 5. Membrane transport; 6. Signal transduction; 7. Signal molecules and interactions; 8. Folding sorting and degradation; 9. Replication and repair; 10. Transcription; 11. Translation; 12. Amino acid metabolism; 13. Biosynthesis of other secondary metabolites; 14. Carbohydrate metabolism; 15. Energy metabolism; 16. Glycan biosynthesis and metabolism; 17. Lipid metabolism; 18. Metabolism of cofactors and vitamins; 19. Metabolism of other amino acids; 20. Metabolism of terpenoids and polyketides; 21. Nucleotide metabolism; 22. Overview; 23. Xenobiotics biodegradation and metabolism.

图 5 多裂骆驼蓬 KEGG 通路分类统计图

Fig. 5 Statistical diagram of KEGG pathway of *Peganum multisectum*

这四大类可细分为 23 个小类, 其中翻译和信号传导获得注释信息最多。将注释基因映射到药物合成相关通路, 发现有 240 个 Unigene 与药物合成通路相关, 参与合成的次生代谢产物类型有萜类、苯丙素类、黄酮类、生物碱类等, 这为后续多裂骆驼蓬药用活性成分的研究提供了基础数据。此外, 在多裂骆驼蓬转录组中还发现多个脱落酸和茉莉酸的合成相关基因, 甘草(张春荣等, 2015)的干旱胁迫研究表明此类基因与抗干旱胁迫作用相关, 推测此类基因与多裂骆驼蓬适应西北干旱环境的特性有关。

多裂骆驼蓬转录组 SSR 的发生频率为 4.32%,



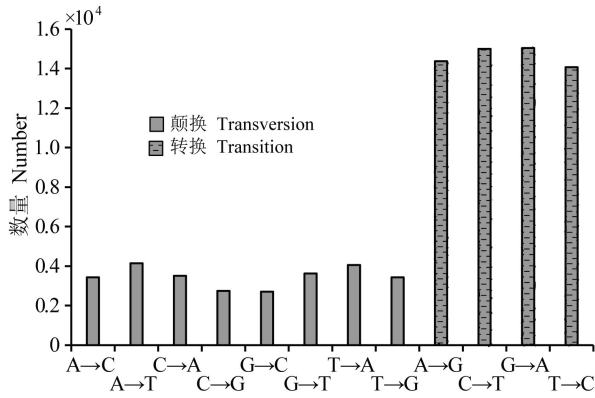


图 6 多裂骆驼蓬 SNP 突变类型统计图

Fig. 6 Statistical diagram of SNP mutation of *Peganum multisectum*

与检索条件相同(即不包含单核苷酸 SSR)的其他物种相比,高于马尾松(*Pinus massoniana*) (2.22%) (杜明凤和丁贵杰,2018)、冷蒿(*Artemisia frigida*) (2.61%) (岳春江等,2016)、杜仲(*Eucommia ulmoides*) (2.91%) (黄海燕等,2013)和杉木(*Cunninghamia lanceolata*) (3.16%) (吴夏雷等,2018),与红松(*P. koraiensis*) (4.24%) (张振等,2015)和灯盏花(*Erigeron breviscapus*) (4.79%) (陈茵等,2014)相差较小,明显低于半夏(*Pinellia ternata*) (16.24%) (王森等,2014)和刺梨(*Rosa roxburghii*) (20.37%) (鄢秀芹等,2015)。这种差异可能是与物种选择、组装方法或筛选软件的不同有关。此外,该物种转录组 SSR 的优势基元为三核苷酸重复,这与马尾松(杜明凤和丁贵杰,2018)、杉木(吴夏雷等,2018)、刺梨(鄢秀芹等,2015)等研究结果相一致。有研究表明,转录区的三核苷酸重复在面对自然选择时表现出积极响应的作用,且该基序是在编码区受到重大突变压力时而存在的一种丰富的核苷酸重复基序(李彦等,2018),即当植物表现出某些抗逆性时三核苷酸重复分布较为丰富,多裂骆驼蓬具有较强的耐干旱能力进一步证实了该结论。此外,我们推测,随着干旱环境的胁迫,该物种产生了相应的抵御和适应机制,并逐渐形成了丰富的三核苷酸重复结构。通过分析和挖掘多裂骆驼蓬转录组 SSR 信息,可为今后该物种 SSR 分子标记的开发及其遗传多样性研究提供生物信息学基础。

本研究首次获得了多裂骆驼蓬的转录组数据,通过生物信息学的方法对该数据进行分析,内容涵盖基因比对、CDS 预测、通路注释、SNP 检测和 SSR 检测,对该植物的基因、通路以及分子标记等方面有了初步的认识,弥补了本植物在分子数据方面的空缺,也进一步丰富了白刺科植物的转录组数据库,同时也为多裂骆驼蓬的物种分化研究、谱系地理学研究、遗传多样性研究和系统发育学研究提供了数据保障,为本种植物资源的开发利用奠定了基础。

## 参考文献:

- ANGIOSPERM PHYLOGENY GROUP, CHASE MW, CHRISTENHUSZ MJM, et al., 2016. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV [J]. *Bot J Linn Soc*, 181(1): 1-20.
- CHEN Y, LI CT, JIANG NH, et al., 2014. SSR information in *Erigeron breviscapus* transcriptome and polymorphism analysis [J]. *Chin J Chin Mat Med*, 39(7): 1220-1224. [陈茵,李翠婷,姜倪皓,等,2014.灯盏花转录组中 SSR 位点信息分析及其多态性研究[J].*中国中药杂志*, 39(7): 1220-1224]
- CHEN Y, LI K, ZHANG HL, et al., 2015. Diversity of bacterial communities in rhizosphere soil of *Peganum harmala* L. in desert steppe [J]. *Medical Inform*, 28(4): 380-381. [陈豫,李凯,张洪亮,2015.中药骆驼蓬抗肿瘤作用的研究进展[J].*医学信息*, 28(4): 380-381]
- CHENG L, CHEN JX, LI YL, et al., 2018. Diversity of bacterial communities in rhizosphere soil of *Peganum harmala* L. in desert steppe [J]. *Arid Zone Res*, 35(4): 977-983. [程琳,陈吉祥,李彦林,等,2018.荒漠草原植物骆驼蓬根际土壤细菌群落分析[J].*干旱区研究*, 35(4): 977-983.]
- DU MF, DING GJ, et al., 2018. Analysis of SSR loci of functional gene linked to drought resistance based on transcriptome sequences in *Pinus massoniana* under drought stress [J]. *Sci Silv Sin*, 31(5): 12-22. [杜明凤,丁贵杰,2018.基于马尾松干旱转录组的抗旱功能 SSR 位点分析[J].*林业科学研究*, 31(5): 12-22.]
- DUAN JA, ZHOU RH, ZHAO SX, et al., 1998. Studies on the chemical constituents of *Peganum multisectum* Maxim I. The alkaloids from seeds and antitumour activity [J]. *J China Pharm Univ*, 29(1): 21-23. [段金殿,周荣汉,赵守训,等,1998.多裂骆驼蓬化学成分研究: I.种子生物碱类成分及其抗肿瘤活性[J].*中国药科大学学报*, 29(1): 21-23.]
- FAN ZR, YAO XS, 1998. Constituents and pharmacological effects of *Peganum* [J]. *J Shenyang Pharm Univ*, 9(2): 144-151. [樊峥嵘,姚新生,1992.骆驼蓬属(*Peganum*)植物成

- 分及药理作用 [J]. 沈阳药科大学学报, 9(2): 144-151.]
- FU SH, LEI M, ZHANG YQ, et al., 2019. Mining and bioinformatic analysis of *FPPS* gene from *Dysphania schraderiana* transcriptome database [J]. *Guihaia*, 39(6): 831-842. [付苏宏, 雷鸣, 张勇群, 等, 2019. 菊叶香藜转录组数据库中 *FPPS* 基因的挖掘与生物信息学分析 [J]. 广西植物, 39(6): 831-842.]
- GRABHERR M, YASSOUR M, LEVIN J, et al., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome [J]. *Nat Biotechnol*, 29(7): 644.
- HUANG HY, DU HY, WUYUN TN, et al., 2013. Development of SSR molecular markers based on transcriptome sequencing of *Eucommia ulmoides* [J]. *Sci Silv Sin*, 49(5): 176-181. [黄海燕, 杜红岩, 乌云塔娜, 等, 2013. 基于杜仲转录组序列的 SSR 分子标记的开发 [J]. 林业科学, 49(5): 176-181.]
- HUANG X, JIN J, YU J, et al., 2017. Transcriptome sequencing of *Panax pseudoginseng* and identification of key enzyme genes in triterpenoids saponin biosynthesis [J]. *Genom Appl Biol*, 36(6): 2531-2538. [黄勋, 金健, 余洁, 等, 2017. 羽叶三七的转录组测序与三萜皂苷生物合成的关键酶基因的识别 [J]. 基因组学与应用生物学, 36(6): 2531-2538.]
- JIANG XM, WU YF, XIAO FM, et al., 2014. Transcriptome analysis for leaves of five chemical types in *Cinnamomum camphora* [J]. *Hereditas*, 36(1): 58-68. [江香梅, 伍艳芳, 肖复明, 等, 2014. 樟树 5 种化学类型叶片转录组分析 [J]. 遗传, 36(1): 58-68.]
- LI K, XUE XQ, ZHANG HL, 20015. Research progress on extraction technology and anti-tumor mechanism of effective components from *Peganum harmala* [J]. *Xinjiang J Trad Chin Med*, 33(2): 80-82. [李凯, 薛小青, 张洪亮, 2015. 骆驼蓬有效成分的提取工艺及抗肿瘤机制研究进展 [J]. 新疆中医药, 33(2): 80-82.]
- LIU JX, 2003. Effects of extracts of *P. multisectum* Maxim Bobr on seed germination and seedling growth and physiological character of oat [J]. *J Triticeae Crops*, 36(6): 117-119. [刘建新, 2003. 多裂骆驼蓬提取物对燕麦种子萌发和幼苗生长及生理特性的影响 [J]. 麦类作物学报, 23(4): 117-119.]
- LIU JX, HU HB, WANG X, 2008. Study on allelopathy of aqueous extract from *Peganum multisectum* (Maxim.) Bobr. on *Perennial Ryegrass* (*Lolium perenne* L.) and its physiological-biochemical manifestation [J]. *Acta Agr Sin*, 16(4): 374-379. [刘建新, 胡浩斌, 王鑫, 2008. 多裂骆驼蓬水浸液对多年生黑麦草的化感作用与生理生化表现 [J]. 草地学报, 16(4): 374-379.]
- LIU JX, WANG X, LI BP, 2011. Effects of exogenous nitric oxide donor on photosynthesis and xanthophyll cycle of *Peganum multisectum* seedlings under NaCl stress [J]. *J Desert Res*, 31(1): 137-141. [刘建新, 王鑫, 李博萍, 2011. 外源一氧化氮供体对盐胁迫下多裂骆驼蓬幼苗光合作用和叶黄素循环的影响 [J]. 中国沙漠, 31(1): 137-141.]
- LIU YX, 1995. A study on origin and for mation of the chinese desert floras [J]. *J Syst Evol*, 33(2): 131-141. [刘嫵心, 1995. 试论我国沙漠地区植物区系的发生与形成 [J]. 植物分类学报, 33(2): 131-141.]
- MA J, WANG XL, 1998. the species and distribution of genus *Peganum* L. in the desert area of China [J]. *J Desert Res*, 18(2): 131-136. [马骥, 王勋陵, 1998. 中国荒漠地区骆驼蓬属植物种类与分布 [J]. 中国沙漠, 18(2): 131-136.]
- MENG WJ, ZHANG AL, JIANG LX, et al., 2019. Cloning and bioinformatics analysis of *Psammosilene tunicoides* transcription factor *ptMYC2* [J]. *Guihaia*, 39(10): 1350-1358. [孟文俊, 张爱丽, 蒋乐晓, 等, 2019. 金铁锁转录因子 *ptMYC2* 的克隆和生物信息学分析 [J]. 广西植物, 39(10): 1350-1358.]
- MORIYA Y, ITOH M, OKUDA S, et al., 2007. KAAS: an automatic genome annotation and pathway reconstruction server [J]. *Nucl Acids Res*, 35(Suppl. 2): W182-W185.
- PATEL S, SHAH D, PANCHAL H, 2014. *De Novo* RNA seq assembly and annotation of *Trigonella foenum-graecum* L. (SRR066197) [J]. *Legume Genom Genet*, 5(7) 1-7.
- QUINLAN A, HALL I, 2010. BEDTools: a flexible suite of utilities for comparing genomic features [J]. *Bioinformatics*, 26(6): 841-842.
- SCHAFLEITNER R, KUMAR S, LIN CY, et al., 2013. The okra (*Abelmoschus esculentus*) transcriptome as a source for gene sequence information and molecular markers for diversity analysis [J]. *Gene*, 517(1): 27-36.
- SHENDURE J, JI H, 2008. Next-generation DNA sequencing [J]. *Nat Biotechnol*, 26(10): 1135.
- THIEL T, MICHALEK W, VARSHNEY R, et al., 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.) [J]. *Theor Appl Genet*, 106(3): 411-422.
- UNIPROT C, 2018. UniProt: A worldwide hub of protein knowledge [J]. *Nucl Acids Res*, 47(D1): D506-D515.
- WANGS, ZHANG Z, JIANG NH, et al., 2014. SSR information in transcriptome of *Pinellia ternata* [J]. *J Chin Med Mat*, 37(9): 1566-1569. [王森, 张震, 姜倪皓, 等, 2014. 半夏转录组中的 SSR 位点信息分析 [J]. 中药材, 37(9): 1566-1569.]
- WU Q, DUAN XQ, CHEN X, et al., 2012. EST-SSRs markers of *Taxus chinensis* based on high throughput sequencing [J]. *Chin J Chin Mat*, 37(24): 3728-3733. [吴琼, 段小群, 陈旭, 等, 2012. 基于高通量测序的红豆杉 EST-SSRs 标记研究 [J]. 中国中药杂志, 37(24): 3728-3733.]
- WU XL, DONG L, SUN YH, et al., 2018. Development of EST-SSR molecular marker based on sequences of *Cunninghamia lanceolata* transcripts under drought stress [J]. *J NE For Univ*, 46(2): 1-5. [吴夏雷, 董黎, 孙宇涵, 等, 2018. 基于干旱胁迫下杉木转录组序列的 EST-SSR 分子标记开发 [J]. 东北林业大学学报, 46(2): 1-5.]
- XU LR, HUANG CJ, 1993. *Flora Reipublicae Popularis Sinicae* [M]. Beijing: Science Press, 43(1): 123-126. [徐朗然, 黄成就, 1993. 中国植物志 [M]. 北京: 科学出版社, 43(1): 123-126.]

- XU XP, BU R, CHEN XH, 2008. Investigation on resource distribution and evaluation of medicinal prospect of *Peganum harmala* in Northwest China [M]. J Shaanxi Coll Trad Chin Med, 31(6): 70-71. [徐小平, 卜瑞, 陈兴华, 2008. 我省及西北地区骆驼蓬资源分布状况调查与药用前景评估 [J]. 陕西中医学院学报, 31(6): 70-71.]
- XUE LG, ZHAO GL, WANG YM, 2005. Studies on the inhibition and disinsection of fungistatic agent extracted from *Peganum multisectum* Bobr. [J]. Microbiology, 32(1): 48-51. [薛林贵, 赵国林, 王毅民, 2005. 多裂骆驼蓬提取液的抑菌杀虫作用研究 [J]. 微生物学通报, 32(1): 48-51.]
- YAN XQ, LIU M, AN HM, 2015. Analysis on SSR information in transcriptome and development of molecular markers in *Rosa roxburghii* [J]. Acta Horti Sin, 42(2): 341-349. [鄢秀芹, 鲁敏, 安华明, 2015. 刺梨转录组 SSR 信息分析及其分子标记开发 [J]. 园艺学报, 42(2): 341-349.]
- YANG YF, LIU HW, QIU DY, 2017. Transcriptome characterization of *Stellera chamaejasme* with illumina sequencing technology [J]. Chin Trad Herb Drugs, 48(22): 4740-4747. [杨艳芳, 刘洪伟, 邱德有, 2017. 基于高通量测序的瑞香狼毒转录组数据分析 [J]. 中草药, 48(22): 4740-4747.]
- YE XZ, LIU D, LUO JJ, et al., 2019. Transcriptome analysis for rare and endangered plants of *Semiliquidambar cathayensis* [J]. Bull Bot Res, 39(2): 276-286. [叶兴状, 刘丹, 罗佳佳, 等, 2019. 濒危珍稀植物半枫荷的转录组分析 [J]. 植物研究, 39(2): 276-286.]
- YUAN C, PENG F, YANG ZM, et al., 2017. EST-SSR identification, markers development of *Ligusticum chuanxiong* based on *Ligusticum chuanxiong* transcriptome sequences [J]. Chin J Chin Mat, 42(17): 3332-3340. [袁灿, 彭芳, 杨泽茂, 等, 2017. 川芎转录组 SSR 分析与 EST-SSR 标记的开发 [J]. 中国中药杂志, 42(17): 3332-3340.]
- YUE CJ, CHEN CC, GUO FX, et al., 2016. Data mining of simple sequence repeats in transcriptome sequences of Mongolia medicinal plant *Artemisia frigida* Willd [J]. J Agric Sci Technol, 18(6): 31-43. [岳春江, 陈川川, 郭凤仙, 等, 2016. 蒙药冷蒿转录组 SSR 信息分析 [J]. 中国农业科技导报, 18(6): 31-43.]
- ZHAN GZ, ZHANG HG, MO C, et al., 2015. Transcriptome sequencing analysis and development of EST-SSR markers for *Pinus koraiensis* [J]. Sci Silv Sin, 51(8): 114-120. [张振, 张含国, 莫迟, 等, 2015. 红松转录组 SSR 分析及 EST-SSR 标记开发 [J]. 林业科学, 51(8): 114-120.]
- ZHANG CL, QIN ZJ, WANG GZ, et al., 2012. Transcriptome and RNA-seq technology [J]. Biotechnol Bull, (12): 51-56. [张春兰, 秦孜娟, 王桂芝, 等, 2012. 转录组与 RNA-Seq 技术 [J]. 生物技术通报, (12): 51-56.]
- ZHANG CR, SANG XY, QU M, et al., 2015. *De novo* sequencing and analysis of root transcriptome to reveal regulation of gene expression by moderate drought stress in *Glycyrrhiza uralensis* [J]. Chin J Chin Mat, 40(24): 4817-4823. [张春荣, 桑雪雨, 渠萌, 等, 2015. 基于转录组测序揭示适度干旱胁迫对甘草根基因表达的调控 [J]. 中国中药杂志, 40(24): 4817-4823.]
- ZHANG DY, WEN H, WANG W, et al., 2017. Transcriptional analysis of terpenoid biosynthesis in *Aconitum carmichaelii* [J]. Chin J Exp Trad Med Form, 32(16): 45-50. [张大燕, 文欢, 王伟, 等, 2017. 乌头萜类生物合成代谢的转录组学分析 [J]. 中国实验方剂学杂志, 32(16): 45-50.]
- ZHANG FQ, GAO QB, KHAN G, et al., 2014. Comparative transcriptome analysis of aboveground and underground tissues of *Rhodiola algida*, an important ethno-medicinal herb endemic to the Qinghai-Tibetan Plateau. [J]. Gene, 553(2): 90-97.
- ZHANG SP, HONG JJ, QIU SL, et al., 2016. Sequencing and analysis of the transcriptome of *Gynura bicolor* [J]. Acta Horti Sin, 43(5): 935-946. [张少平, 洪建基, 邱珊莲, 等, 2016. 紫背天葵高通量转录组测序分析 [J]. 园艺学报, 43(5): 935-946.]
- ZHAO T, WANG ZT, BRANFORD-WHITE CJ, et al., 2011. Classification and differentiation of the genus *Peganum* indigenous to China based on chloroplast *trnL-F* and *psbA-trnH* sequences and seed coat morphology. [J]. Plant Biol, 13(6): 940-947.
- ZHOU H, ZHANG X, LIU TY, et al., 2012. Data processing and gene discovery of high-throughput transcriptome sequencing [J]. Jiangxi Sci, 30(5): 607-611. [周华, 张新, 刘腾云, 等, 2012. 高通量转录组测序的数据分析与基因发掘 [J]. 江西科学, 30(5): 607-611.]
- ZOU LQ, KUANG XJ, LI Y, et al., 2016. *De novo* sequencing and analysis of root transcriptome to reveal regulation of gene expression by moderate drought stress in *Glycyrrhiza uralensis* [J]. Chin J Chin Mat, 41(22): 4138-4143. [邹丽秋, 匡雪君, 李滢, 等, 2016. 人参属药用植物转录组研究进展 [J]. 中国中药杂志, 41(22): 4138-4143.]

(责任编辑 蒋巧媛)