

橡胶草幼叶与成熟叶片转录组比较分析

谢全亮, 杨起航, 刘坤杰, 刘 乐, 谢双全, 王 斐, 李鸿彬*

石河子大学农学院和生命科学学院, 新疆 石河子
Email: xiequanliang001@shzu.edu.cn, *lihbab@163.com

收稿日期: 2021年1月18日; 录用日期: 2021年3月1日; 发布日期: 2021年3月11日

摘 要

橡胶草(*Taraxacum kok-saghyz* L. Rodin)是天然橡胶产业最有发展潜力的新型产胶替代作物。除产胶之外它还能够产生菊粉等大量的次生代谢产物。为了进一步阐明橡胶草叶片中次生代谢产物合成的分子机制, 本研究以新疆橡胶草幼叶和成熟叶片为材料, 采用华大BGISEQ-500平台进行转录组测序, 分别得到165,935,798条和166,460,570条的clean reads, 并注释到Nr、KOG、KEGG、InterPro和Swiss-Prot数据库中。Unigene共有58,402条被注释结果, 通过差异基因表达分析、GO分类和KEGG代谢通路分析, 代谢通路分属于136类, 其中包括倍萜类化合物、核糖体、黄酮醇和花青素等次生代谢物质生物合成, 最后进行差异基因的蛋白互作分析。该研究结果对橡胶草叶片药用活性成分生物的合成与代谢、关键酶基因的克隆以及分子标记开发等研究有一定的参考价值。

关键词

橡胶草, 转录组, 差异基因, 功能分析, 次生代谢产物

Comparative Transcriptome Analysis of Young and Mature Leaves of *Taraxacum kok-saghyz* L. Rodin

Quanliang Xie, Qihang Yang, Kunjie Liu, Le Liu, Shuangquan Xie, Fei Wang, Hongbin Li*

College of Agricultural Sciences and College of Life Science, Shihezi University, Shihezi Xinjiang
Email: xiequanliang001@shzu.edu.cn, *lihbab@163.com

Received: Jan. 18th, 2021; accepted: Mar. 1st, 2021; published: Mar. 11th, 2021

*通讯作者。

文章引用: 谢全亮, 杨起航, 刘坤杰, 刘乐, 谢双全, 王斐, 李鸿彬. 橡胶草幼叶与成熟叶片转录组比较分析[J]. 植物学研究, 2021, 10(2): 85-96. DOI: 10.12677/br.2021.102013

Abstract

Rubber grass (*Taraxacum kok-saghyz* L. Rodin) is the most promising rubber-producing alternative crop for new natural rubber. It can biosynthesize large amounts of secondary metabolites such as inulin. In order to further clarify the molecular mechanism of secondary metabolite biosynthesis in rubber grass leaves, this study used young and mature leaves of Xinjiang rubber grass as materials, and used the BGISEQ-500 platform for transcriptome sequencing to obtain 165,935,798 and 166,460,570 clean reads, respectively. Annotate to Nr, KOG, KEGG, InterPro and Swiss-Prot databases. Unigene has a total of 58,402 annotated results, and then performed differential gene expression analysis, GO classification, and KEGG metabolic pathway analysis. The metabolic pathways are classified into 136 categories, including the synthesis of sesquiterpenoids, ribosomes, flavonols, anthocyanins, etc., biosynthesis of secondary metabolites, and finally protein-protein interaction analysis of differential genes. The results of this study have certain reference value for the research on the biosynthesis and metabolism of medicinal active ingredients in rubber grass leaves, the cloning of key enzyme genes, and the development of molecular markers.

Keywords

Taraxacum kok-saghyz L. Rodin, Transcriptome, Differential Gene, Functional Analysis, Secondary Metabolite

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

橡胶草(*Taraxacum kok-saghyz* L. Rodin, TKS)也称为俄罗斯蒲公英,属于菊科植物,蒲公英属,橡胶草种,原产于新疆天山边界附近的特克斯河盆地与哈萨克斯坦的天山河谷接壤处[1]。染色体数为 $2N = 2X = 16$ 条[2][3],二倍体的多年生草本植物[4]。橡胶草的基因组大小约1.29 Gb,共包含46,000多个基因编码蛋白质[5]。TKS的根中含有天然橡胶(Natural rubber, NR),其中NR占总干重的3%~28%(多年生)[6][7],并且橡胶分子大小和性能与巴西橡胶树(*Hevea brasiliensis*)中的天然橡胶极为相似[8][9]。由于近缘物种短角蒲公英 *Taraxacum brevicorniculatum* (TB)与TKS在生境和形态上较为类似且也能产天然橡胶[10][11]。TKS有着能够在寒冷和温带地区广泛生长等特点,生命期相对较短,含胶量和质量高,易转基因和收获方法较简单等优点[12][13],橡胶草已被作为研究天然橡胶生物合成的理想模式植物,也正在成为天然橡胶替代生产最有发展潜力的作物。TKS根部除了含NR外,还含有约28%的次生代谢产物菊粉,菊粉是食品工业和生物乙醇制造中重要原料[12][13],其副产品产值更大。然而,橡胶草在成为商业上可行的作物之前,仍然需要在橡胶产量和农艺性状等方面的改善。

早期由于橡胶草基础数据的缺乏,以基因组学、转录组学和蛋白质组学等研究方法,挖掘橡胶草农艺性状相关的关键基因越来越受到关注[14]。Luo等利用转录组挖掘橡胶产量相关的SNPs标记,为解析橡胶合成相关遗传性状和辅助选择育种基础[15]。Cao等利用MeJA诱导橡胶草根的转录组分析发现,与橡胶形成相关的TFs与NR合成基因之间相互作用,共同调节次生代谢产物并影响橡胶的生物合成[16]。Nowicki等利用基因组微卫星基因座的研究,gSSR和eSSR标记结果表明,TKS具有低至中等的遗传多样性,而该物种缺乏种群结构,这为菊科相关生物的多样性分析提供了一种具有分子标记功能性的工具,

揭示了 TKS 遗传多样性与菊科物种密切相关[17]。对 TKS 根部进行丙酮萃取, 根据其提取物中三萜类化合物的组成分析, 发现氧化角鲨烯环化酶 TkOSC1 和 TkCYP716A263 在橡胶草的乳胶中过表达后, 会产生大量被修饰的五环三萜类化合物[18]。橡胶草乳胶中敲除 OSC 基因, 可以减少根中天然橡胶中三萜的含量。而在 RNAi-OSC 植株中并没有表现出明显的表型变化, 含胶量也未受影响。这为 TKS 育种计划以及开发具有稳定低三萜烯含量的橡胶草植物提供坚实的基础[19]。在橡胶草中使用 CRISPR/Cas9 敲除 TkRALFL1 基因后的植株根系生物量更高, 菊粉和天然橡胶的产量也更高[20]。橡胶草乳胶中敲除 TkCPTL1 基因, 致使合成顺式-1,4-异戊二烯(IPP)减少, 但 TKS 根中三萜和菊粉的含量却升高了, 而三醇的含量未受影响。对 TkCPTL1-RNAi 植物的胶乳进行分析发现, 具有与橡胶粒子相似的粒子将三萜烯包裹在由小橡胶粒子蛋白 SRPP 和稳定的磷脂壳中。相反, TkCPTL2 基因表达下调, 没有改变橡胶草的表型, 这表明 TkCPTL2 蛋白在橡胶草中的功能有待进一步探究[21]。TkCPTL1-RNAi 植物和野生型橡胶草乳胶蛋白质组的比较发现了假定蛋白, 这些蛋白与类异戊二烯生物合成途径, 代谢产物以及合成橡胶的橡胶粒子膜的成分有关[21]。Xie 等对橡胶草成熟根的全蛋白鉴定建立了高质量的数据集, 并提出橡胶延伸因子(REF)和 SRPP 在橡胶合成过程中同等重要[22]。

Benninghausl 等利用蛋白质组学和代谢组学的方法, 在产胶和非产胶的橡胶草根中, Branase 诱导橡胶 RNA 降解, 对 TKS 根部代谢产物和蛋白质组成的影响, 并鉴定出多种乳胶特异性蛋白, 还包括迄今为止尚未在植物中进行讨论的类脂蛋白。Barnase 的表达使有胶橡胶草显出非乳胶的表型, 此结果有望作为植物与环境进一步相互作用过程中, 对植物产乳胶进行前瞻性分析最有价值的基础, 以阐明整个植物界中乳胶的进化和具体分布。此外, 不同代谢途径的分子相互作用, 也反应了橡胶合成网络关系有了更深入的了解[23]。Zhang 等利用间接超声结合尺寸排阻色谱法改进橡胶草橡胶提取结果显示, 比常规提取方法所需的提取时间更短, 效率更高, 且该方法不影响 TKS 的橡胶分子量分布[24]。Panara 等比较非胶、低含胶量(Low Rubber, LR)和高含胶量(High Rubber, HR)三种橡胶草转录组差异基因表达, 发现 HR 胶乳中橡胶含量较高, 是由于许多直接参与橡胶合成基因表达呈正相关的结果, 并表明天然橡胶的产生在转录水平上受到高度调控。而 LR 中较低胶含量可能与天然橡胶生物合成竞争的其它次生代谢产物合成中涉及的基因的较高表达有关[25]。

橡胶草的叶片内含有一定量的花青素和菊糖, 叶片的副产物在饲料、食品营养和药理等领域有着潜在的应用价值[8] [26]。并且橡胶草叶片中的菊苣酸有潜在的抗氧化作用, 可作为酚类化合物的抗氧化来源, 也是中草药和生产健康食品/饲料的营养成分[27]。综上所述发现, 目前关于新疆橡胶草幼嫩与成熟叶片转录组的比较分析研究的鲜有报道, 在本研究中, 对新疆橡胶草幼叶及成熟叶片的转录组测序后的序列进行了组装、Unigene 功能注释、参考基因组、差异表达基因分析、并在功能上注释差异表达基因分类和次级代谢产物合成分析等研究。橡胶草的转录数据也扩大并丰富了可利用的橡胶草基因组资源, 并可能加速橡胶草的种质资源和育种改良。

2. 材料与方法

2.1. 材料与试剂

新疆橡胶草幼嫩叶片与成熟叶片取自石河子大学生命科学学院, 新疆植物医学资源与利用教育部重点实验室种质资源圃内种植的新疆橡胶草。分别取种植 30 d 后的幼苗期(leaf_1)和种植 120 d 后的成熟期(leaf_2)长势良好, 将健康的橡胶草叶片组织, 迅速置于液氮中并保存至-80℃冰箱, 以备试验所用。

2.2. 橡胶草叶片总 RNA 提取

使用改良的 CTAB 法提取橡胶草叶片总 RNA, 在 CTAB 方法的基础上, 加 1%的 PVPP, 加入 β -巯

基乙醇的 SL 500 μL ，两次氯仿抽提，并用 RNA 级的酚:氯仿:异戊醇 = 25:24:1 进一步抽提，最后晾干沉淀，滴加 RNase-Free ddH₂O 30 μL ，得到提取好的 RNA 溶液，标记备份-80 $^{\circ}\text{C}$ 保存。

2.3. 橡胶草叶片转录组测序

以新疆橡胶草幼苗期与成熟期种植苗的叶片为材料，分别提取叶片总 RNA 后，使用 Nanodrop (Agilent 2100 Bioanalyzer)检测总 RNA 的浓度、总量、OD260/280、OD260/230 和 RIN 值，RNA 纯度使用 NanoDropTM 的紫外分光光度计检测，结果为 A 级。将合格总 RNA 片段化后，逆转录为 cDNA，在 cDNA 的 5'-和-3' 末端加上接头，再进行 PCR 扩增，将扩增产物变性环化，最终获得单链环状的 DNA 文库。上机 BGISEQ-500 进行转录组测序，测序于华大基因公司(深圳)完成。

2.4. 原始测序数据过滤

测序的原始数据包含接头污染、低质量以及未知碱基 N 含量(>5%)过高的 reads，数据分析之前需要去除这些 reads 后得到 clean reads，以保证结果的可靠性。得到 clean reads 之后，使用 HISAT 将 clean reads 比对到参考基因组序列，使用 Bowtie 2 将 clean reads 比对到参考序列，之后再使用 RSEM 计算基因和转录本的表达水平。

2.5. 差异表达基因检测及生物信息学分析

根据样品组之间的差异表达基因(DEG)，使用 DESeq 2 和 PossionDis 方法进行差异基因检测，为了更直观地展示每个样品在不同 FPKM 区间的基因数目，对 FPKM (FPKM ≤ 1 、 $1 < \text{FPKM} < 10$ 、FPKM ≥ 10) 的三种情况进行了基因数目的统计，同时使用 MA plot 展示 DEG 的分布。然后根据差异基因检测结果，对其 Gene Ontology (GO)功能以及 KEGG (Kyoto Encyclopedia of Genes and Genomes)生物通路分类富集分析。

2.6. 差异表达基因的蛋白互作分析

为了研究 leaf_1 和 leaf_2 优先积累的蛋白质网络，使用了在线数据库 STRING (<https://string-db.org/>)进行了差异表达基因的蛋白互作分析(Protein-protein interaction, PPI)。通过 PPI 分析，具有相互作用的 DEG 通常具有相似的功能。根据 STRING 蛋白互作数据库，对每组差异表达基因进行蛋白互作分析。

3. 结果与分析

3.1. 转录组测序与组装分析

对新疆橡胶草幼嫩和成熟叶片分别进行转录组测序，两个组织分别获得了 214.16 Mb 和 211.67 Mb 的 Raw reads 量，经过过滤后 reads 质量，最终获得幼嫩叶片 165.94 Mb 和成熟叶片 166.46 Mb 的 clean reads (表 1)。

Table 1. Quality statistics of filtered reads

表 1. 过滤后的 reads 质量统计

样品名 Sample	Total Raw Reads (Mb)	Total Clean Reads (Mb)	Total Clean Bases (Gb)	Clean Reads Q20 (%)	Clean Reads Q30 (%)	Clean Reads Ratio (%)
leaf_1	214.16	165.94	16.59	96.36	88.96	77.48
leaf_2	211.67	166.46	16.65	96.36	88.83	78.64

Total Raw Reads (Mb): 过滤前的 reads 数; Total Clean Reads (Mb): 过滤后的 reads 数; Total Clean Bases (Gb): 过滤后的碱基总数; Clean Reads Q20 (%): 过滤后的 reads 中质量值大于 20 的碱基数占总碱基数的百分比; Clean Reads Q30 (%): 过滤后的 reads 中质量值大于 30 的碱基数占总碱基数的百分比; Clean Reads Ratio (%): 过滤后的 reads 的比例。

Table 1. Reference genome comparison result statistics**续表 1.** 参考基因组比对结果统计

样品名 Sample ^a	总 reads 数 Total Clean Reads ^b	总基因组覆盖率 Total Mapping Ratio ^c	单基因覆盖率 Uniquely Mapping Ratio ^d
leaf_1	165,935,798	74.73%	37.76%
leaf_2	166,460,570	67.66%	33.63%

^aSample: 样品名; ^bTotal Clean Reads: 过滤后的 reads 总数; ^cTotal Mapping Ratio: 比对上参考基因组的 clean reads 比例; ^dUniquely Mapping Ratio: 唯一比对上参考基因组某一位置的 clean reads 比例。

过滤后的 reads 中质量值大于 20 和 30 的碱基数分别占总碱基数的百分比的 96.36%和 88.90%。过滤后的 reads 的比例占 78%。转录本的质量指标测定, 其 GC 值分别为 40.81%和 40.76%。

3.2. 橡胶草叶片转录组与参考基因组的比对分析

测序结果可用于后进行参考基因组对比, 共获得 332,396,368 条 clean reads, 其中种植 30 d 幼苗叶片(leaf_1)共有 165,935,798 条, 种植 120 d 后的成熟叶片(leaf_2)共有 166,460,570 条, 与对比参考基因组分别占 74.73%和 67.66%。在 Uniquely 匹配时其中匹配率分别为 37.76%和 33.63%。参考基因组中总基因数为 47,643 条基因, 转录测定发现有新的转录本分别为(leaf_1) 35,663 条和(leaf_2) 34,119 条, 还预测了新基因数分别为 14,890 条和 14,009 条(表 2)。

Table 2. Statistics of the number of genes and transcripts**表 2.** 基因、转录本数目统计表

样品名 Sample	基因总数 Total Genes	已知基因数 Known Genes	新基因数 Novel Genes	转录本总数 Total Transcripts	已知转录本数 Known Transcripts	新转录本数 Novel Transcripts
leaf_1	49,820	34,930	14,890	66,082	30,419	35,663
leaf_2	49,297	35,288	14,009	65,187	31,068	34,119

使用 Trinity 对 clean reads 进行组装并获得转录本(Transcript), leaf_1 与 leaf_2 的 Transcript 总数分别为 66,082 个和 65,187 个, 平均长度分别为 737 nt 和 785 nt, 其中 200~1000 nt 范围的转录本分别有 83,461 个和 94,831 个, 占总量的 75.65%和 72.98%; 1000~2000 nt 的 Transcript 分别有 19,547 个和 25,608 个, 占总量的 17.72%和 19.71%; 大于 2000 nt 的 Transcript 分别有 7316 个和 9506 个, 占总量的 6.63%和 7.32%。

对转录本进行聚类去冗余得到 leaf_1 和 leaf_2 的 Unigene, leaf_1 与 leaf_2 Unigene 的总数分别为 66,749 个和 76,456 个, 平均长度分别为 917 nt 和 955 nt, 其中 200~1000 nt 的 Unigene 分别有 44,108 个和 48,959 个, 占总量的 66.08%和 63.04%; 1000~2000 nt 的 Unigene 分别有 16,193 个和 19,677 个, 占总量的 24.26%和 25.74%; 大于 2000 nt 的 Unigene 分别有 6448 个和 7820 个, 占总量的 9.66%和 10.23%。GC (%)含量分别为 40.92%和 40.74%。本实验对转录组数据的组装质量、Transcript 和 Unigene 的长度分布情况进行了统计和分析(表 1, 表 2), 结果发现为 300~1000 nt 长度的 Unigene 所占比重较大, 说明测序质量较高可用于后续 Unigene 的功能注释。

3.3. 橡胶草叶片转录组 All-Unigene 的 CDS 预测

对新疆橡胶草叶片的所有 Unigene 中的 CDS 通过进行 Blast 比对 Swiss-Prot 数据库和 Hmmscan, 搜索 Pfam 蛋白同源序列, 从而预测编码区域。从序列长度分布中我们共获得了 55,120 条 CDS 序列片段(图 1), 总长度为 52,588,077 nt, 其中 100~1000 nt 的有 36,043 条占 65.40%, 1000~2000 nt 的有 14,799 条占 26.85%, 2000~3000 nt 的有 3110 条占 5.64%, 大于 3000 nt 的有 1164 条占 2.11%。

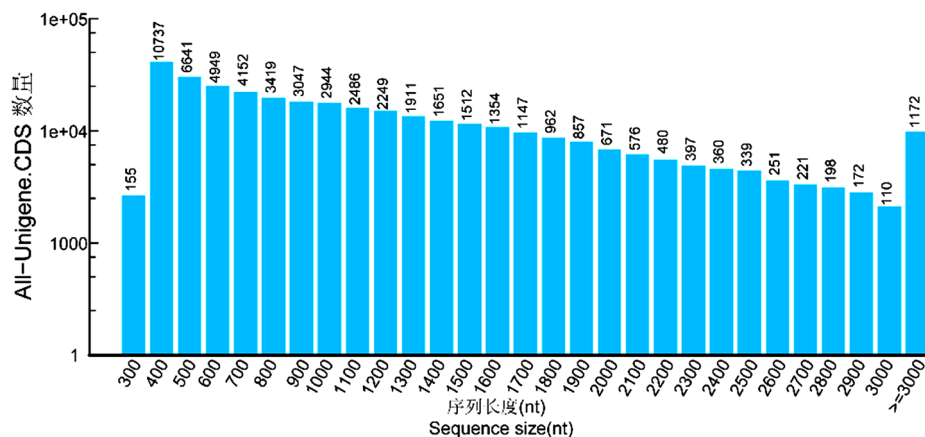


Figure 1. CDS data length distribution for transcriptome of *T. koksghz* leaves

图 1. 橡胶草叶片转录组 CDS 数据长度分布

3.4. 橡胶草叶片转录组 Unigene 功能注释

对测序后组装得到的橡胶草叶片转录组 All-Unigene 分别注释到七大功能数据库 NR、NT、GO、Swiss-Prot、InterPro、KOG、KEGG 的数据库(图 2(A)), 对每个数据库注释的 Unigene 数目进行统计, 共有 99,253 条 Unigene 有对应的功能信息, 其中, 在 NR 中有 58,402 条, 占总的 58.84%; NT 有 29,578 条, 占总的 29.80%; GO 有 19,809 条, 占总的 19.96%; Swiss-Prot 有 38,614 条, 占总的 38.90%; InterPro 有 47,437 条, 占总的 47.79%; KOG 有 45,893 条, 占总的 46.24%; KEGG 有 43,383 条, 占总的 43.71%。

橡胶草叶片转录组 Unigene 的 NR 功能分类

已获得橡胶草叶片组装的 All-Unigene 注释到 NR 数据库, 统计共有 58,402 条 Unigene 被注释结果。在菜蓟(*Cynara cardunculus* var. *scolymus*)、葡萄(*Vitis vinifera*)、芝麻(*Sesamum indicum*)、中果咖啡(*Coffea canephora*)和其它物种中都有同源序列分布。其中, Unigene 与菜蓟相似序列占有 61.35%; 与葡萄相似序列占 4.06%; 与芝麻相似序列占有 2.42%; 与中粒咖啡相似的则有 2.08%; 还有 30.09%的 Unigene 属于其他序列, 可能还包含了橡胶草自身特有的基因序列(图 2(B))。

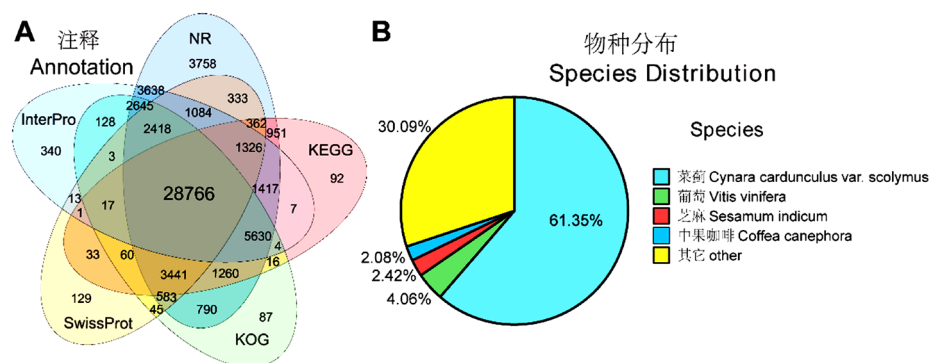


Figure 2. All-Unigene species distribution in the database. A: All-Unigene annotated to the seven functional databases including NR, NT, GO, Swiss-Prot, InterPro, KOG, KEGG database and functional classification, respectively. B: Distribution of unique gene sequences of All-Unigene and cardoon, grape, sesame, medium grain coffee and *T. koksghz*

图 2. All-Unigene 在数据库中的物种分布。A: All-Unigene 分别注释到七大功能数据库 NR、NT、GO、Swiss-Prot、InterPro、KOG、KEGG 的数据库和功能分类。B: All-Unigene 与菜蓟、葡萄、芝麻、中粒咖啡和橡胶草自身特有的基因序列分布情况

3.5. 橡胶草叶片差异基因表达量分析

将橡胶草叶片 leaf_1 与 leaf_2 进行差异基因表达基因数目对比统计, 发现 leaf_1 与 leaf_2 对比中 12,637 个表达基因上调, 8103 个基因表达下调, 33,509 个基因没有差异表达(图 3)。

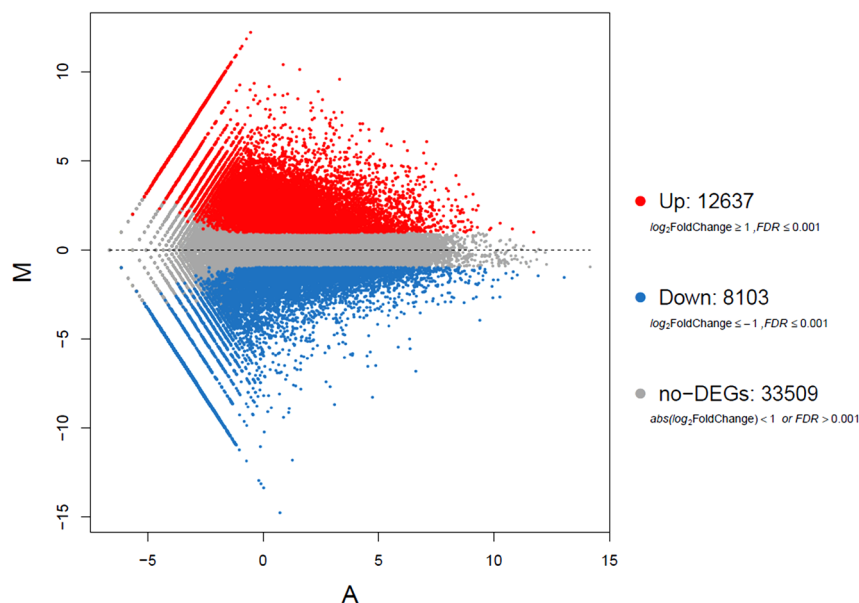


Figure 3. Statistical map of differentially expressed genes

图 3. 差异表达基因数量统计图

3.6. 橡胶草叶片差异基因的功能注释

为进一步了解橡胶草幼叶片及成熟叶片响应的相关分子机制, 对 Differentially expressed gene (DEGs) 进行 Gene Ontology (GO)富集分析。GO 主要分为生物过程(biological process)、分子功能(molecular function)和细胞组分(cellular component)三大类别。leaf_1 与 leaf_2 的差异基因表达经 GO 分析分别分布的三大类别中, 共有 52 个类型, 其中生物过程中包括代谢过程(2750)、细胞过程(2381)和单一生物过程(1525); 细胞组成包括细胞部分(2067)、膜组成(1445)和组织部分(1264); 分子功能包括催化活性(3131)、结合(2538)和运输活性(2538)等。另外, 大多数差异基因与生物过程相关, 而参与分子功能的差异基因相对较少。在细胞组分中, 以参与细胞、细胞组分、膜、膜组分、细胞器构成的富集最明显。参与生物过程的差异基因主要富集在代谢过程和细胞过程这两个类别; 而分子功能中催化活性和结合是 leaf_1 和 leaf_2 分子功能差异基因富集的主要类别(图 4)。

3.7. 差异表达基因的 Pathway 功能分析

为了进一步研究差异基因的生物学功能, 进行了 KEGG 通路分析。分析差异基因 KEGG 代谢通路结果显示, 差异表达基因能映射到新陈代谢(Metabolism)、遗传信息处理(Genetic Information Processing)、环境信息处理(Environmental Information Processing)、细胞进程(Cellular Processes)和有机系统(Organismal Systems)的 5 大类 19 个一级通路主要途径, 其中涉及代谢通路的基因最多, 共有 10,434 个, 占到注释 KEGG 中的基因比例的 61.78%。其次是遗传信息处理, 占 KEGG 中的基因比例的 23.23% (图 5)。一级通路中包括翻译过程有 777 个基因, 碳水化合物代谢有 1572 个基因, 折叠基因 1375 个, 环境适应 662 个基因, 整体代谢和概观代谢基因 3761 个, 氨基酸代谢 948 个基因, 转运和分解代谢基因 786 个, 脂质代

谢基因 724 个等(图 5)。这些结果表明, 橡胶草幼嫩与成熟叶片差异表达基因主要与折叠与降解, 翻译和碳水化合物代谢有关。

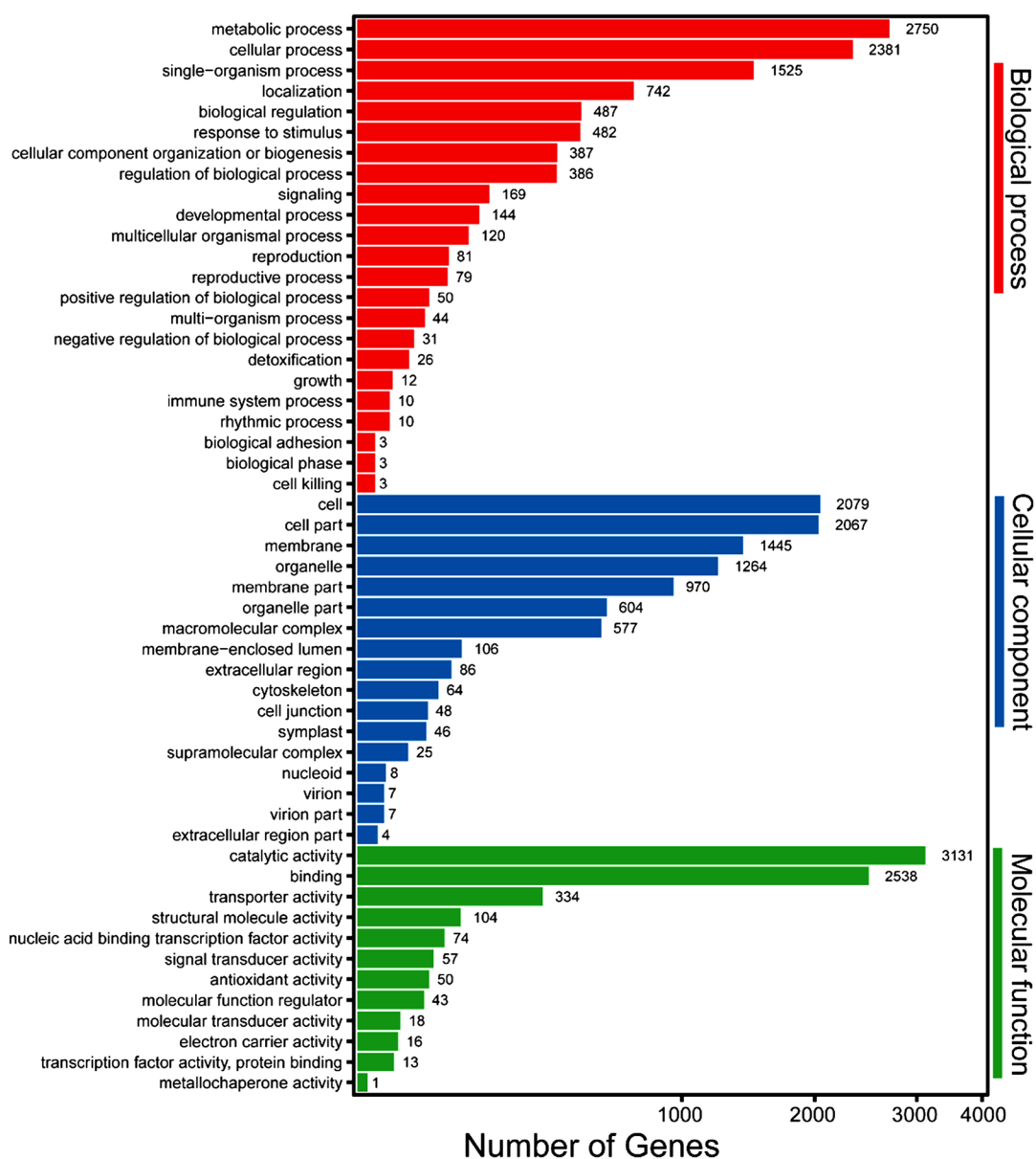


Figure 4. GO functional classification of the leaf_1-VS-leaf_2 differentially expressed genes

图 4. Leaf_1-VS-leaf_2 差异基因 GO 功能分类图

除此之外, 还分析了前 20 个 KEGG 亚通路(图 5(B)), 结果显示, 其中包含 3525 种基因的最丰富的途径是碳代谢。第二个包含 2032 个具有次级代谢产物合成的基因。第三子途径具有用于氨基酸的生物合成的 675 种基因, 其次是用于嘧啶和嘌呤代谢的 642 和 641 种基因, 以及内质网蛋白合成中的 579 种基因。这些 KEGG 亚途径中还富含涉及苯丙素和蔗糖代谢, 糖酵解, 苯丙素类生物合成以及氨基糖和核苷酸糖代谢。这些基因对于许多植物, 特别是次级代谢产物和天然橡胶生产植物中类异戊二烯的生物合成至关重要。

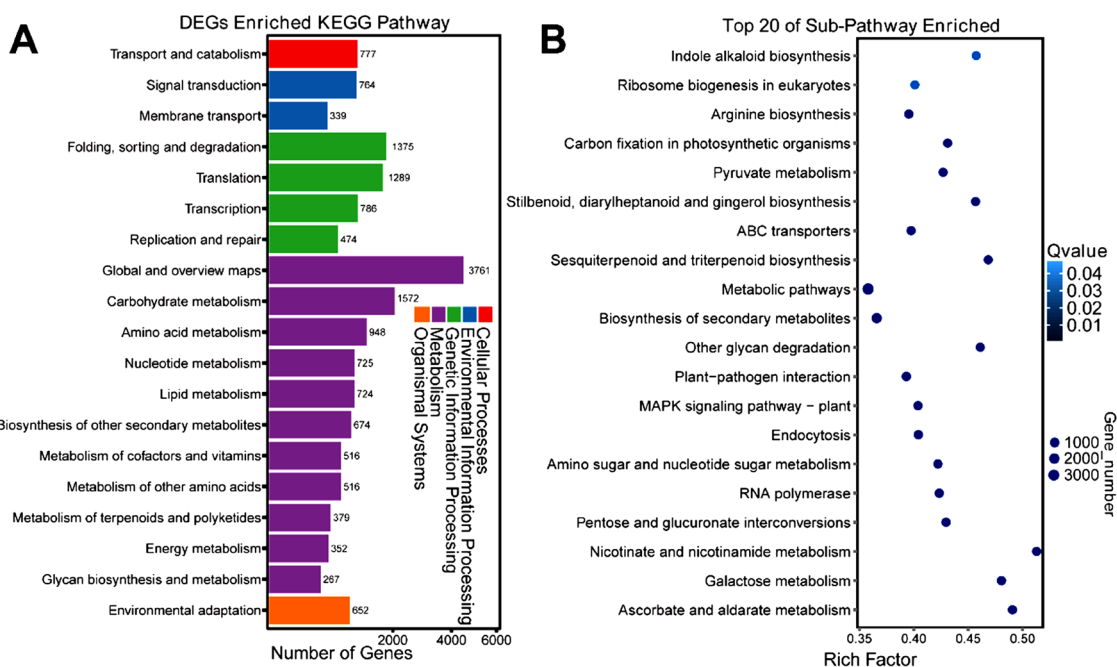


Figure 5. KEGG Pathway classification chart of differentially expressed genes. A, the X-axis represents the proportion of the gene, the Y-axis represents the KEGG functional classification. B, top 20 of sub-pathway enriched. The X-axis represents the enrichment factor value (Rich Factor), the Y-axis represents the path name. The color of the dots in the figure represents the Q-value, and the lighter the color, the greater the value, the deeper the smaller, the latter represents the more significant the enrichment results; the size of the dots represents the number of DEGs, and the larger the number represents, the smaller the number represents.

图 5. 差异基因 KEGG Pathway 分类。A, leaf_1-VS-leaf_2 轴表示基因所占比例, Y 轴表示 KEGG 功能分类。B, 前 20 个亚 KEGG 通路富集。X 轴代表富集因子值(Rich Factor), Y 轴代表通路名称, 图中圆点的颜色代表 Q-value, 颜色越浅值越大, 越深值越小, 值越小代表富集。结果越显著; 圆点的大小代表 DEGs 数目, 点越大代表数目越多, 越小代表数目越少

3.8. 差异表达基因的蛋白互作分析

橡胶草叶片的蛋白质-蛋白质相互作用(PPI)网络分析可以反映蛋白质数据集的关键要素。为了研究橡胶草幼嫩叶片 leaf_1 和成熟叶片 leaf_2 优先积累的蛋白质的网络, 使用了在线数据库 STRING。对于 leaf_1 和 leaf_2 优先积累的蛋白质, 可以观察到 18 个网络簇, 其中 8 个网络簇由不同的颜色代表上调与下调, 不同大小表示关系紧密程(图 6)。其中最大的簇中的这些基因的蛋白质主要由油菜素类固醇相关蛋白组成。尤其是小亚单位核糖体蛋白和大亚单位核糖体蛋白, 以及天冬氨酸氨基甲酰转移酶具有相互作用网络的大多数关键节点。

4. 讨论

早在 1931 年, 人们就已经意识到俄罗斯蒲公英具有产胶、食品及其它副产物等应用价值[14]。橡胶草叶内含次生代谢产物和花青素, 能够改善关节柔韧性, 增强血管强度和抑制炎症等药用价值[9] [28]。随着基因组、转录组、蛋白组和代谢组测序技术不断更新完善, 组学研究为挖掘相关调控机制关键基因/蛋白提供了重要技术方法。其中近期快速发展的转录组测序(RNA-Seq)是用于研究功能基因组学分析技术。目前橡胶草尤其是原产地新疆橡胶草叶片的转录组信息较少, 利用茉莉酸刺激橡胶草产胶转录组分析发现, IPP 是三萜和橡胶合成的关键中间物质, 三萜类物质和橡胶是通过 IPP 的转换互相合成[16] [23]。比较 LR 和 HR 橡胶草根转录组, 也说明次生代谢产物与橡胶合成呈负相关[25]。

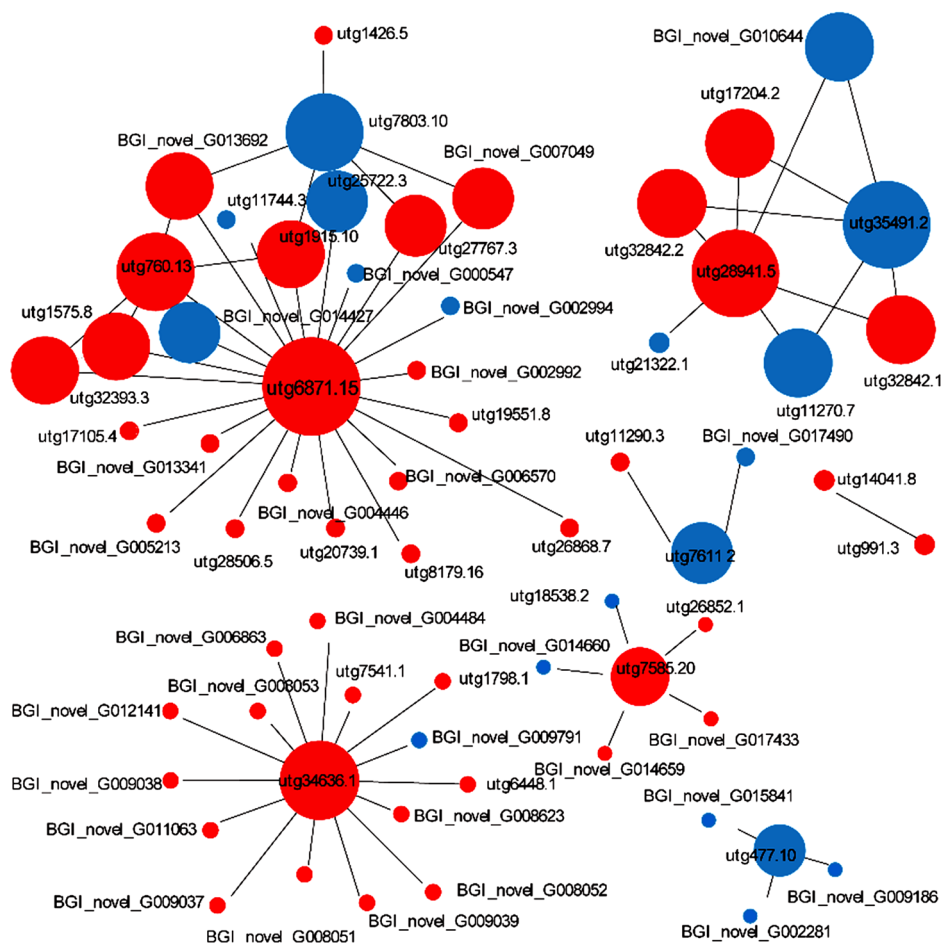


Figure 6. Network diagram of proteins interaction. Red represents up-regulated proteins, blue represents down-regulated proteins, and the size of the circle represents the number of interaction relationships. The larger the circle, the denser the relationship

图 6. 蛋白互作网络图。红色表示上调，蓝色表示下调，圆圈大小表示相互作用的关系的个数，圆的大小表示表关系的强度，圆圈越大表示关系越密集

AtPAP1 在短角蒲公英中的异源表达导致营养组织显示出红色/紫色色素沉着表型，并积累了大量的花色苷。该表型与在拟南芥中过表达 *AtPAP1* 所描述的表型一致，并且与在烟草，甘蓝型油菜和番茄中的异源表达相类似[29] [30] [31]。次生代谢产物是植物中药用活性成分的重要组成部分[32]，通过构建橡胶草不同组织器官的转录组文库，能更全面地研究 TKS 的次生代谢产物生物合成途径中关键基因的具体功能机制，定向提升次生代谢产物要用价值[33]。但次生代谢产物与橡胶合成的具体调控机制未明确阐述。由于橡胶草转基因相关研究数据缺乏，橡胶合成关键基因与橡胶合成精确调控还需多方面验证。本研究通过 RNA-Seq 测序共获得了 33,205 Mb 的 clean reads，将其与参考基因组对比后，匹配率 70% 以上的基因在基因组中匹配到，再通过幼嫩叶片与成熟叶片差异基因表达分析，并注释到 KEGG 的代谢通路中，其中花青素与萜类生物合成相关较多。蛋白互作预测分析油菜素类固醇合成核糖体蛋白合成途径成为主要互作网络关键点。成功地建立了 TKS 叶片转录组高质量的数据，以精确地区分幼嫩叶片和成熟叶片在转录水平的理论分析，获取更多的转录组数据将有助于从数据集中挖掘的一些实例，将为通过生物技术手段，提高橡胶草叶片次生代谢产物的产量和有效成分含量以及活性成分的合成途径以及更深入地了解橡胶草叶片的生理作用和生物学理论研究提供理论基础。

基金项目

国家自然科学基金地区基金项目(32060072); 石河子大学高层次人才科研启动项目(XJ2020001202)。

参考文献

- [1] Cornish, K. (2001) Biochemistry of Natural Rubber, a Vital Raw Material, Emphasizing Biosynthetic Rate, Molecular Weight and Compartmentalization, in Evolutionarily Divergent Plant Species. *Natural Product Reports*, **18**, 182-189. <https://doi.org/10.1039/a902191d>
- [2] 罗士苇. 橡胶草: 产胶植物的介绍之一[J]. 科学通报, 1950, 1(8): 559-564.
- [3] Volis, S., Uteulin, K. and Mills, D. (2009) Russian Dandelion (*Taraxacum Kok-saghyz*): One More Example of Over Collecting in the Past? *Journal of Applied Botany and Food Quality*, **83**, 60-63.
- [4] Krotkov, G. (1945) A Review of Literature on *Taraxacum kok-saghyz* Rod. *The Botanical Review*, **11**, 417-461. <https://doi.org/10.1007/BF02861139>
- [5] Lin, T., Xu, X., Ruan, J., et al. (2017) Genome Analysis of *Taraxacum kok-saghyz* Rodin Provides New Insights into Rubber Biosynthesis. *National Science Review*, **5**, 78-87. <https://doi.org/10.1093/nsr/nwx101>
- [6] Cornish, K., Kopicky, S.L., McNulty, S.K., et al. (2016) Temporal Diversity of *Taraxacum koksaghyz* Plants Reveals High Rubber Yield Phenotypes. *Biodiversitas*, **17**, 847-856. <https://doi.org/10.13057/biodiv/d170262>
- [7] Cornish, K. (2017) Alternative Natural Rubber Crops: Why Should We Care? *Technologies & Innovations*, **18**, 245-256. <https://doi.org/10.21300/18.4.2017.245>
- [8] Hahn, T., Klemm, A., Ziesse, P., et al. (2016) Optimization and Scale-Up of Inulin Extraction from *Taraxacum kok-saghyz* Roots. *Natural Product Communications*, **11**, 689-692. <https://doi.org/10.1177/1934578X1601100535>
- [9] Qiu, J., Sun, S., Luo, S., et al. (2014) Arabidopsis AtPAP1 Transcription Factor Induces Anthocyanin Production in Transgenic *Taraxacum brevicorniculatum*. *Plant Cell Reports*, **33**, 669-680. <https://doi.org/10.1007/s00299-014-1585-8>
- [10] Musto, S., Barbera, V., Maggio, M., et al. (2016) Crystallinity and Crystalline Phase Orientation of Poly(1,4-cis-isoprene) from *Hevea brasiliensis* and *Taraxacum kok-saghyz*. *Polymers for Advanced Technologies*, **27**, 1082-1090. <https://doi.org/10.1002/pat.3774>
- [11] Bach, T.J. (1995) Some New Aspects of Isoprenoid Biosynthesis in Plants. In: Kader, J.C. and Mazliak, P., Eds., *Plant Lipid Metabolism*, Springer, Dordrecht, 321-334. https://doi.org/10.1007/978-94-015-8394-7_90
- [12] 罗士苇, 冯午, 吴相钰. 橡胶草的研究: 部分 I. 新疆产橡胶草的形态观察[J]. 中国科学, 1951, 2(3): 373-379.
- [13] Iaffaldano, B., Cardina, J. and Cornish, K. (2018) Hybridization Potential between the Rubber Dandelion *Taraxacum kok-saghyz* and Common Dandelion *Taraxacum officinale*. *Ecosphere*, **9**, e02115. <https://doi.org/10.1002/ecs2.2115>
- [14] 谢全亮, 李鸿彬, 王旭初. 橡胶草 90 年来主要研究成果及最新研究进展[J]. 植物科学学报, 2019, 37(3): 404-412.
- [15] Luo, Z., Iaffaldano, B.J., Zhuang, X., et al. (2017) Analysis of the First *Taraxacum kok-saghyz* Transcriptome Reveals Potential Rubber Yield Related SNPs. *Scientific Reports*, **7**, Article No. 9939. <https://doi.org/10.1038/s41598-017-09034-2>
- [16] Cao, X.W., Yan, J., Lei, J.L., et al. (2017) De Novo Transcriptome Sequencing of Meja-Induced *Taraxacum koksaghyz* Rodin to Identify Genes Related to Rubber Formation. *Scientific Reports*, **7**, Article No. 15697. <https://doi.org/10.1038/s41598-017-14890-z>
- [17] Nowicki, M., Zhao, Y., Boggess, S.L., et al. (2019) *Taraxacum kok-saghyz* (Rubber Dandelion) Genomic Microsatellite Loci Reveal Modest Genetic Diversity and Cross-Amplify Broadly to Related Species. *Scientific Reports*, **9**, Article No. 1915. <https://doi.org/10.1038/s41598-019-38532-8>
- [18] Pütter, K.M., van Deenen, N., Müller, B., et al. (2019) The OSC1 Enzymes and CYP716A263 Produce a High Variety of Triterpenoids in the Latex of *Taraxacum koksaghyz*. *Scientific Reports*, **9**, Article No. 5942. <https://doi.org/10.1038/s41598-019-42381-w>
- [19] van Deenen, N., Unland, K., Prüfer, D., et al. (2019) Oxidosqualene Cyclase Knock-Down in Latex of *Taraxacum koksaghyz* Reduces Triterpenes in Roots and Separated Natural Rubber. *Molecules*, **24**, 2703. <https://doi.org/10.3390/molecules24152703>
- [20] Wieghaus, A., Pütter, K.M. and Gronover, C.S. (2019) Loss of Function Mutation of the Rapid Alkalinization Factor (RALF1)-Like Peptide in the Dandelion *Taraxacum koksaghyz* Entails a High-Biomass Taproot Phenotype. *PLoS ONE*, **14**, e0217454. <https://doi.org/10.1371/journal.pone.0217454>
- [21] Niephaus, E. and Muller, B. (2019) Uncovering Mechanisms of Rubber Biosynthesis in *Taraxacum koksaghyz*—Role

- of Cis-Prenyltransferase-Like 1 Protein. *The Plant Journal*, **100**, 591-609. <https://doi.org/10.1111/tpj.14471>
- [22] Xie, Q.L., Ding, G.H., Zhu, L.P., *et al.* (2019) Proteomic Landscape of the Mature Roots in a Rubber-Producing Grass *Taraxacum Kok-saghyz*. *International Journal of Molecular Sciences*, **20**, 2596. <https://doi.org/10.3390/ijms20102596>
- [23] Benninghaus, V.A., Deenen, N.V., Müller, B., *et al.* (2019) Comparative Proteome and Metabolome Analyses of Latex-Exuding and -Non-Exuding *Taraxacum kok-saghyz* Roots Provide Insights into Laticifer Biology. *Journal Experimental Botany*, **19**, erz512. <https://doi.org/10.1093/jxb/erz512>
- [24] Zhang, N., Guo, T.Y., Ma, X., *et al.* (2019) Rational Rubber Extraction and Simultaneous Determination of Rubber Content and Molecular Weight Distribution in *Taraxacum kok-saghyz* Rodin by Size-Exclusion Chromatography. *Chromatographia*, **82**, 1459-1466. <https://doi.org/10.1007/s10337-019-03773-2>
- [25] Panara, F., Lopez, L., Daddiego, L., *et al.* (2019) Comparative Transcriptomics between High and Low Rubber Producing *Taraxacum koksaghyz* R. Plants. *BMC Genomics*, **19**, Article No. 875. <https://doi.org/10.1186/s12864-018-5287-4>
- [26] 仇键, 张继川, 罗世巧, 等. 橡胶草的研究进展[J]. 植物学报, 2015, 50(1): 133-141.
- [27] Molinua, M.G., Piluzza, G., Campesib, G., *et al.* (2019) Antioxidant Sources from Leaves of Russian Dandelion. *Chemistry & Biodiversity*, **16**, e1900250. <https://doi.org/10.1002/cbdv.201900250>
- [28] 梁素钰, 王文帆, 刘滨凡, 等. 能源橡胶草的综合利用研究[J]. 能源研究与信息, 2010, 26(4): 219-236.
- [29] Zuluaga, D.L., Gonzali, S., Loreti, E., *et al.* (2008) *Arabidopsis thaliana* MYB75/PAP1 Transcription Factor Induces Anthocyanin Production in Transgenic Tomato Plants. *Functional Plant Biology*, **35**, 606-618. <https://doi.org/10.1071/FP08021>
- [30] Li, X., Gao, M.J., Pan, H.Y., *et al.* (2010) Purple Canola: *Arabidopsis* PAP1 Increases Antioxidants and Phenolics in *Brassica napus* Leaves. *Journal of Agricultural and Food Chemistry*, **58**, 1639-1645. <https://doi.org/10.1021/jf903527y>
- [31] Zhou, L.L., Zeng, H.N., Shi, M.Z., *et al.* (2008) Development of Tobacco Callus Cultures over Expressing Arabidopsis PAP1/MYB75 Transcription Factor and Characterization of Anthocyanin Biosynthesis. *Planta*, **229**, 37-51. <https://doi.org/10.1007/s00425-008-0809-y>
- [32] 吴琼, 孙超, 陈士林, 等. 转录组学在药用植物研究中的应用[J]. 世界科学技术(中医药现代化), 2010, 12(3): 457-462.
- [33] 王尧龙, 黄璐琦, 袁媛, 等. 药用植物转录组研究进展[J]. 中国中药杂志, 2015, 40(11): 2055-2061.