

## 滇黄精全长转录组测序及生物信息学分析

米琪<sup>1#</sup>, 赵艳丽<sup>1#</sup>, 徐萍<sup>1</sup>, 于梦雯<sup>1</sup>, 张萱<sup>1</sup>, 涂振华<sup>1</sup>, 李春华<sup>3</sup>,  
郑国伟<sup>1\*</sup>, 陈佳<sup>2\*</sup>

(1. 云南中医药大学中药学院, 云南昆明 650500; 2. 云南中医药大学民族医药学院, 云南昆明 650500;  
3. 贡山县农业农村局农业综合发展服务中心, 云南怒江傈僳族自治州 673500)

**摘要:** 本文通过 PacBio Sequel 平台对滇黄精根茎进行全长转录组测序, 并对测序结果进行生物信息学分析, 旨在为滇黄精基因功能的研究提供参考。最终测得 polymerase reads 为 1 120 485 个, 总数据共 77.73 GB。在 NR 数据库中, 有 41 864 个同源序列比对到 5 个物种; 有 40 506 个在 KOG 数据库得到注释, 根据其功能分为 26 类; 有 69 060 个得到 GO 注释, 分为细胞组分、分子功能及生物学过程等 3 大类的 32 个功能组; 有 45 779 个注释到 KEGG 数据库的 145 条代谢途径中, 其中鉴定与多糖合成有关的转录本在糖酵解/糖异生代谢通路中有 127 条, 淀粉和蔗糖代谢通路有 144 条, 氨基酸糖和核苷酸糖代谢通路有 85 条, 果糖和甘露糖代谢通路有 69 条; 此外还检测到 1 781 个转录因子, 分布在 37 个转录因子家族中; 检测到了 180 293 个 SSR 位点, 其中以单碱基重复占比最多, 占有碱基重复类型的 30.48%, 最少的是五碱基重复, 仅为单碱基重复的 0.14%。本研究结果为丰富滇黄精基因组信息, 探索其有效成分合成途径提供重要的数据支持。

**关键词:** 滇黄精; 全长转录组; 种质资源; 多糖合成

中图分类号: R931 文献标识码: A 文章编号: 0513-4870(2024)06-1864-09

## Full-length transcriptome sequencing and bioinformatics analysis of *Polygonatum kingianum*

MI Qi<sup>1#</sup>, ZHAO Yan-li<sup>1#</sup>, XU Ping<sup>1</sup>, YU Meng-wen<sup>1</sup>, ZHANG Xuan<sup>1</sup>, TU Zhen-hua<sup>1</sup>, LI Chun-hua<sup>3</sup>,  
ZHENG Guo-wei<sup>1\*</sup>, CHEN Jia<sup>2\*</sup>

(1. College of Chinese Material Medica, Yunnan University of Chinese Medicine, Kunming 650500, China; 2. College of Ethnic Medicines, Yunnan University of Chinese Medicine, Kunming 650500, China; 3. Gongshan County Agriculture Village Ju Agriculture Comprehensive Development Service Center, Nujiang of the Lisu Autonomous Prefecture, Nujiang 673500, China)

**Abstract:** The purpose of this study was to enrich the genomic information and provide a basis for further development and utilization of *Polygonatum kingianum*. The fresh rhizome of *P. kingianum* was used as the experimental material, and the full-length transcriptome was sequenced by PacBio Sequel platform. The final measured polymerase reads were 1 120 485, with a total of 77.73 GB of data. In NR database, 41 864 homologous sequence alignments were aligned to 5 species; 40 506 were annotated in the KOG database and classified into 26 categories based on their functionality; 69 060 GO annotated 32 functional groups divided into 3 categories: cellular components, molecular functions, and biological processes; 45 779 annotations were added to 145 metabolic

收稿日期: 2023-11-06; 修回日期: 2024-02-06.

基金项目: 国家自然科学基金资助项目 (82360751); 云南中医药大学高层次人才计划 (2019YZG07); 云南青年精英人才项目 (YNWR-QNBJ-2020-289).  
\*共同第一作者.

\*通讯作者 E-mail: gwzhengkm@163.com; 394626580@qq.com

DOI: 10.16438/j.0513-4870.2023-1259

pathways in the KEGG database. Among them, there are 127 identified transcripts related to polysaccharide synthesis in the glycolysis/gluconeogenesis metabolic pathway, 144 in the starch and sucrose metabolic pathway, 85 in the amino acid sugar and nucleotide sugar metabolic pathway, and 69 in the fructose and mannose metabolic pathway. In addition, 1 781 transcription factors were detected, distributed in 37 transcription factor families; 180 293 SSR loci were detected, among which single base repeats accounted for the most, accounting for 30.48% of all base repeats, and at least five base repeats, accounting for only 0.14% of single base repeats. The results of this study provide important data supporting for enriching the genomic information of *P. kingianum* and exploring its effective biosynthetic pathway.

**Key words:** *Polygonatum kingianum*; full-length transcriptome; germplasm resource; polysaccharide synthesis

滇黄精 (*Polygonatum kingianum* Coll. et Hemsl) 为百合科黄精属植物, 与黄精 (*P. sibiricum*) 和多花黄精 (*P. cyrtoneura*) 共同收录于《中华人民共和国药典》(2020版) 中<sup>[1]</sup>, 是云南重要的道地药材之一, 具有补气养阴、健脾润肺等功效, 用于治疗脾胃气虚、体倦乏力等症。主要分布在云南、四川、贵州等西南部省份, 海拔 1 400~2 100 m 的亚热带地区<sup>[2]</sup>。滇黄精根茎中含有大量的糖分、淀粉、蛋白质以及维生素等多种营养成分, 具有保健功能, 久服可以强筋健骨、坚实肌肉, 对人体机能恢复十分有益<sup>[3]</sup>。作为药食同源植物, 早在 2002 年就被卫生部列入《既是食品又是药品的物品名单》, 其保健功效得到国家权威部门认可<sup>[4]</sup>。

滇黄精根茎中富含多糖、甾体皂苷、生物碱、总酚等活性成分以及可溶性糖、蛋白质、还原糖、黄酮等营养成分<sup>[5]</sup>。其中黄精多糖作为中国药典规定的唯一含量测定成分, 具有广泛的生物学功能, 如调节血糖血脂、调节免疫、抗氧化、抗肿瘤、调节免疫等功能<sup>[6]</sup>。多糖的生物合成是一个十分复杂的过程, 滇黄精多糖主要由甘露糖、半乳糖、葡萄糖、阿拉伯糖、木糖、氨基半乳糖、岩藻糖、氨基葡萄糖、鼠李糖、葡萄糖醛酸以及半乳糖醛酸 11 种单糖组成<sup>[7]</sup>。已有研究报道了 20 余种可能参与黄精多糖生物合成的关键酶基因<sup>[8]</sup>, 推动了滇黄精活性成分生物合成的研究。但是目前对于黄精多糖生物合成的研究依然处于前期探索阶段, 对于滇黄精的遗传信息以及多糖、皂苷、营养成分等合成通路研究报道较少。

随着转录组测序技术的发展, 二代转录组测序技术广泛应用于植物研究, 在植物研究中发挥着重要作用。Wang 等<sup>[9]</sup>通过 HiSeq 2500 转录组测序, 确定  $\beta$ -呋喃果糖苷酶 (*INV*)、己糖激酶 (*HK*) 和甘露糖-1-磷酸鸟苷转移酶 (*GMPP*) 是滇黄精多糖代谢途径中酶的关键基因。Wang 等<sup>[10]</sup>利用 Illumina HiSeq 测序平台对黄精全株进行转录组测序, 发现 UDP-葡萄糖基转移酶、 $\beta$ -葡萄糖苷酶等参与了黄精皂苷元生物合成途径。但是二代测序存在读长短, 在建库的时候需要将转录本打

成碎片; 在转录本异构体的组装过程中存在较多的嵌合体错误, 丢失可变剪切等重要信息等缺点, 大大制约了转录组数据的深度利用。三代测序的出现弥补了这一缺点, 相比于二代测序技术边合成边测序<sup>[11]</sup>, 三代测序克服了二代测序在鉴别重复元素等方面的局限性, 有效减少时间, 同时原理上避免了 PCR 的扩增错误, 纠正 GC 偏好性等缺点, 具有更高的通量和测序效率<sup>[12,13]</sup>。本研究采用 PacBio 测序平台对滇黄精进行全长转录组测序, 利用生物信息学软件进行组装和注释, 以期丰富滇黄精转录组数据, 为滇黄精活性成分的生物合成途径奠定基础。

## 材料与方法

**材料** 供试材料为百合科黄精属滇黄精 (*P. kingianum*), 采自云南省红河州石屏县龙武镇黄精种植基地 (23.93 N, 102.41 E, 海拔 2 174.50 m), 快速清洗泥土后取当年新生部分根茎, 45 °C 烘干后测得其多糖含量为 199.80 mg·g<sup>-1</sup>, 其余部分液氮速冻, -80 °C 冰箱保存, 用于后续提取 RNA。

**RNA 提取与检测** 利用琼脂糖凝胶电泳检测样本基因组是否污染; Nanodrop 检测所提 RNA 的纯度; 使用 Nano Drop 和 Agilent 2100 生物分析仪 (Thermo Fisher Scientific, USA) 检测 RNA 的完整性; Qubit 对 RNA 浓度进行精确定量。

**转录组测序文库构建** Oligo (dT) 富集含有 polyA 的 mRNA, 使用 SMARTer PCR cDNA 试剂盒将 mRNA 反转录为 cDNA; PCR 扩增富集成成的 cDNA, 通过循环优化确定 PCR 的最佳条件; 利用磁珠筛选片段进行大规模 PCR, 以获得足够的 cDNA 总量; 全长 cDNA 进行损伤修复、末端修复、链接 SMRT 哑铃型接头, 构建全长转录组文库; 核酸外切酶消化, 去除 cDNA 两端未连接接头的序列; 最后结合引物、绑定 DNA 聚合酶, 形成完整的 SMRT ball 文库。库检合格后, 使用 PacBio SM-RT Iso-seq 平台进行全长转录组测序。测序工作委托武汉迈维生物科技有限公司。

**生物信息分析流程** 使用PacBio SMRTlink对下机后的数据进行处理,首先对单分子测序序列 polymerase reads 进行拆分得到 Subreads,基于同一 polymerase reads 得到的 Subreads 经过自我纠错形成循环共识序列 (circular consensus sequence, CCS),再根据序列是否包含 5'、3'端引物以及 polyA 尾将序列分为全长序列 (full-length non-concatemer, FLNC) 与非全长序列,最后通过 FLNC 聚类,得到一致性转录本系列 (consensus transcripts)。

将去除冗余后的转录本比对到非冗余蛋白序列数据库 (NR, <http://ftp.ncbi.nlm.nih.gov/blast/db>)、SwissProt 蛋白序列数据库 (SwissProt, <http://ftp.ebi.ac.uk/pub/databases/swissprot>)、基因本体数据库 (GO, <http://geneontology.org>)、京都基因与基因组百科全书 (KEGG, <http://www.genome.jp/kegg>) 等数据库,得到转录本的注释结果;对去除冗余后的转录本进行简单重复序列 (simple sequence repeats, SSR) 分析和编码潜能预测得到长链非编码 RNA (long non-coding RNA, Lnc-RNA)。

**多糖合成相关基因挖掘** 基于文献报道的多糖合成通路,根据转录本在 KEGG 中糖酵解/糖异生 (ko00010)、淀粉和蔗糖代谢 (ko00500)、氨基酸糖和核苷酸糖代谢 (ko00520)、果糖和甘露糖代谢 (ko00051) 的注释信息,结合滇黄精多糖含量,通过多糖含量与基因表达量的相关性分析,初步挖掘与滇黄精多糖合成相关的基因和转录本。

**数据分析** 使用 Diamond<sup>[14]</sup> Blastx 软件对转录本进行 KEGG、NR、Swiss-Prot、GO、COG、KOG 功能注释,使用 CNCI<sup>[15]</sup>、CPC2<sup>[16]</sup>、PLEK<sup>[17]</sup> 等编码潜能预测软件进行编码潜能预测,使用 iTAK<sup>[18]</sup> 软件进行转录因子注释,使用 MISA<sup>[19]</sup> 进行 SSR 位点分析;使用 Origin 2019 和迈维云平台进行绘图及 Pearson 相关性分析 ( $P$  value 阈值: 0.05, cor. value 阈值: 0.80)。

## 结果与分析

### 1 全长转录本数据统计

通过提取滇黄精根茎的总 RNA,在 PacBio 平台上建立了 1 个 PacBio ISO-Sep 文库。共获得 77.73 GB 原始数据 (1 120 485 reads),平均长度为 69 375 bp。在 SMRT cell 的测序中得到 CCS bases: 2 295 228 275 bp,平均长度为 2 932 bp,最长序列为 14 832 bp。

插入片段识别处理后,序列可以分成三类:全长序

列、全长非嵌合序列 (FLNC) 和具有 polyA 的全长非嵌合序列。FLNC 用于后续分析,具体数据如表 1 所示。

利用 CD-HIT 软件进行聚类去冗余, total bases: 159 858 002 bp; total number: 55 454; 最短序列: 101 bp; 平均长度: 2 883 bp; 最长序列: 10 346 bp; N50: 3 201 bp。

### 2 基因功能注释和功能分类

使用 Diamond Blastx 软件将滇黄精三代转录本去冗余后得到的转录本序列与 KEGG、NR、Swiss-Prot、GO 以及 KOG 数据库比对,结果显示共有 52 690 (95.00%) 个基因得到注释 (图 1A)。滇黄精转录组在与 NR、GO、KOG、KEGG 和 SwissProt 5 个数据库比对后,均注释到的基因有 29 975 个,单独注释到的基因分别有 2 069、4、5、46 和 0 个 (图 1B)。滇黄精转录组在 NR 数据库中注释得到 52 346 个基因,占比 94.40%,其中与石刁柏 (*Asparagus offioconalis* L.) 的基因配比率最高,有 36 733 个基因,占比为 70.17%,其他基因配比率高的物种分别为油棕 (*Elaeis guineensis* Jacq.)、海枣 (*Phoenix dactylifera* L.)、椰子 (*Cocos nucifera* L.) 和大车前 (*Planago major* L.), 占比分别为 3.77%、3.72%、1.54% 和 0.77% (图 1C)。在 GO 数据库中根据基因功能分为生物学过程、细胞成分和分子功能三大类,其中生物学过程比对到 161 262 个基因,细胞成分比对到 49 755 个基因,分子功能比对到 69 060 个基因 (图 1D)。

将滇黄精转录本比对 KOG 数据库的 25 个功能组进行分类统计,其中滇黄精转录本比对到最多的功能为“一般功能预测”,数目达到 7 841 个,其次为“翻译后修饰、蛋白质转运、伴侣蛋白”,数目达 4 335,最少的为“细胞活性”,仅为 21 (图 2)。

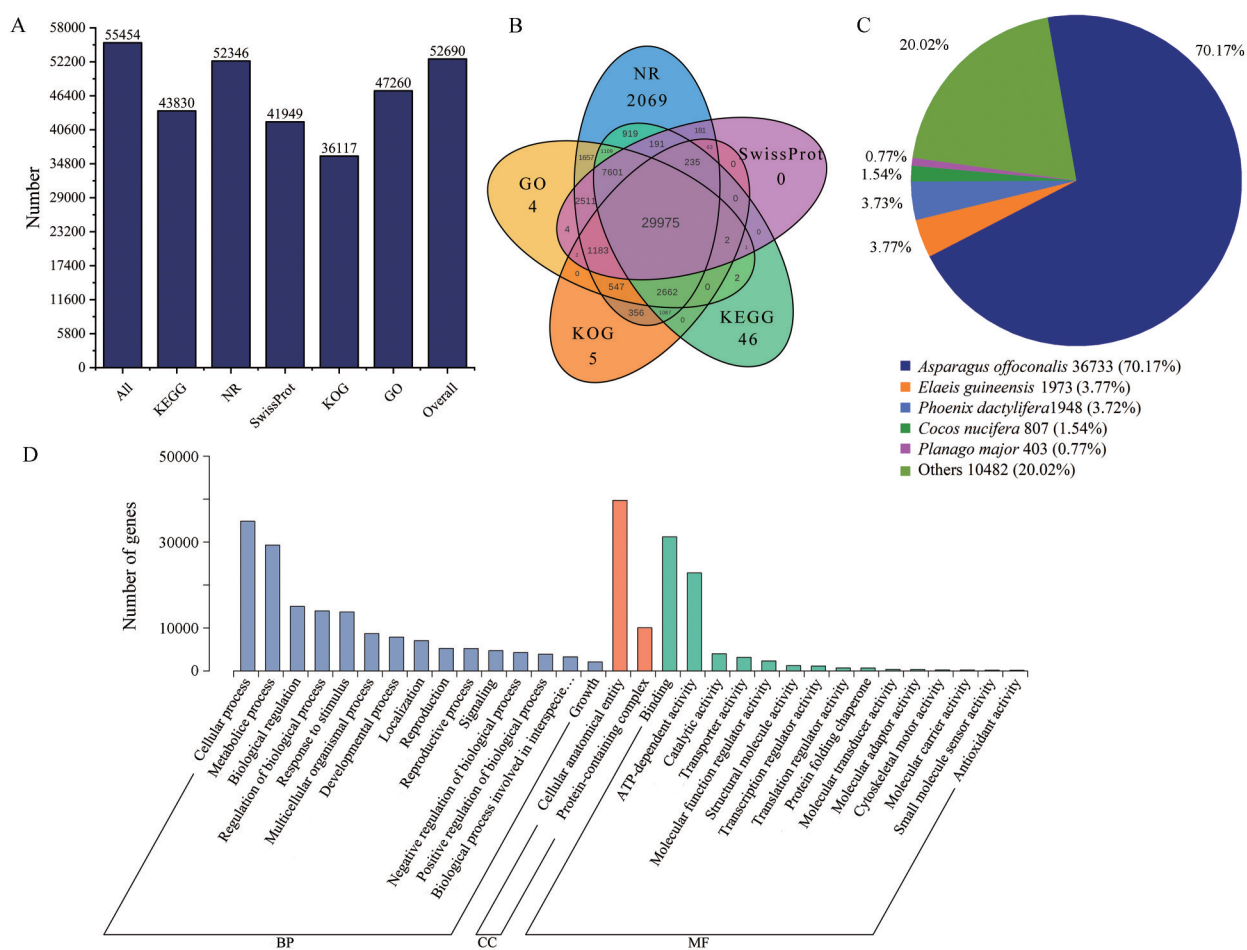
在 KEGG 数据库中,将转录本参与的 KEGG 代谢通路分为 5 个分支:细胞过程、环境信息处理、遗传信息处理、代谢以及有机系统,测得的基因数分别为 1 837、2 880、10 352、29 084 和 1 626 个 (图 3)。

### 3 Lnc-RNA 转录本预测分析

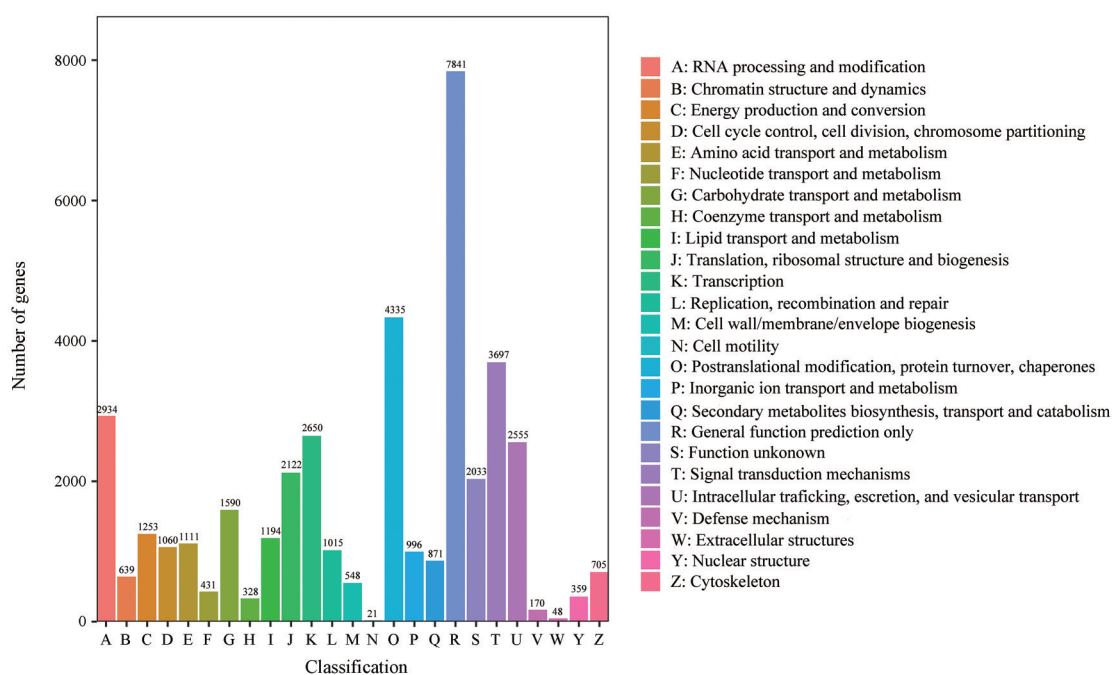
Lnc-RNA 为一类长度大于 200 bp 的长链非编码 RNA。在获得可靠的滇黄精转录本后,将转录本的编码能力进行预测,对 7 大数据库都没有注释到的转录本,根据 Lnc-RNA 的不编码蛋白的功能特点,通过 CNCI、CPC2、PLEK 软件对滇黄精非全长序列进行 Lnc-RNA 预测,取 3 种分析软件鉴定得到的非编码转录本的交集,共得到 1 250 个共有 Lnc-RNA (图 4A)。将 mRNA 与预测出的 Lnc-RNA 进行长度分布密度比

**Table 1** FLNC result statistics. FLNC: Full-length non-concatemer

Sample	FLNC bases/bp	FLNC number	Min number	Average length /bp	Max length /bp	N50 /bp
<i>Polygonatum kingianum</i>	1 953 756 490	678 465	72	2 880	14 758	3 212



**Figure 1** Functional annotation of full-length transcripts of *P. kingianum*. A: Summary of functional annotation results for *P. kingianum* transcripts; B: The Venn diagram between NR, COG, KEGG, Swissprot and GO; C: Species distribution of annotated transcripts in NR database; D: Classification diagram of GO annotation. BP: Biological process; CC: Cellular component; MF: Molecular function



**Figure 2** Classification diagram of KOG annotation

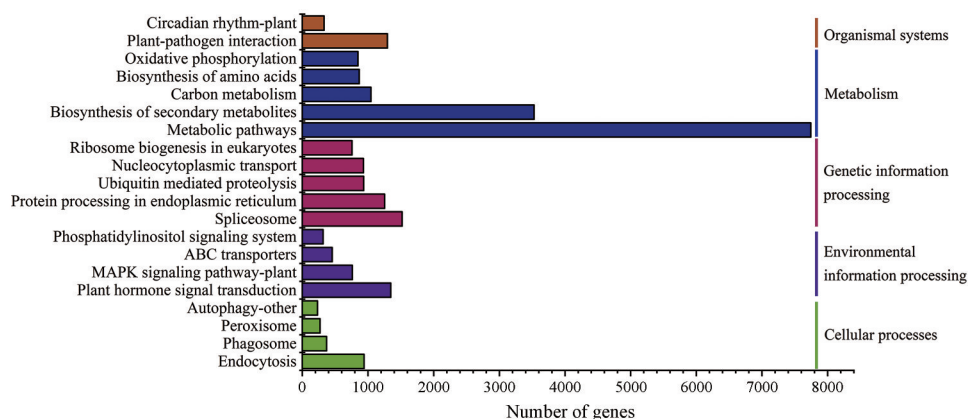


Figure 3 Schematic diagram of KEGG function distribution

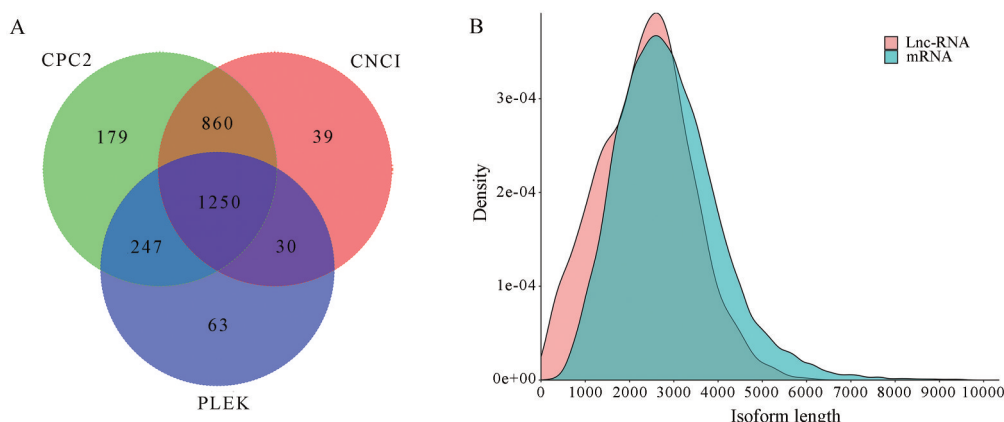


Figure 4 Prediction and analysis of Lnc-RNA transcript of *P. kingianum*. A: Venn plot of Lnc-RNA prediction results of *P. kingianum*; B: Distribution map of Lnc-RNA and mRNA length. CPC2: Contrastive predictive coding; CNCI: Coding-non-coding index; PLEK: Predictor of long non-coding RNAs and messenger RNAs based on an improved k-mer scheme; Lnc-RNA: Long non-coding RNA; mRNA: Messenger RNA

较, 可知在滇黄精非全长序列中, Lnc-RNA 与 mRNA 转录本长度相似, 集中在 0~600 bp (图 4B)。

#### 4 转录因子注释

对滇黄精样本中具有转录因子 (transcription factor, TF) 功能的转录本进行注释, 共鉴定到属于 37 个基因家族的 1 781 个 (图 5)。其中含有转录本数量最多的 10 个家族分别是: C3H (219 条)、FAR1 (164 条)、C2H2 (118 条)、bZIP (97 条)、bHLH (90 条)、MYB (90 条)、NAC (86 条)、SBP (86 条)、TUB (79 条)、GRAS (75 条)。研究显示, C3H、FAR1、C2H2、bZIP 等转录因子在植物生长发育及逆境胁迫应答过程发挥重要作用; MYB 和 bHLH 等转录因子参与了植物次生代谢的调控, 包括萜类、黄酮、木质素、糖类物质的生物合成<sup>[20]</sup>。这些转录因子家族的鉴定为今后滇黄精次生代谢产物的生物合成, 以及逆境响应等提供了条件。

#### 5 SSR 检测及引物设计

5.1 SSR 检测 通过 MISA (<https://webblast.ipk-gatersleben.de/misa/>) 对滇黄精转录本进行 SSR 检测, 共检

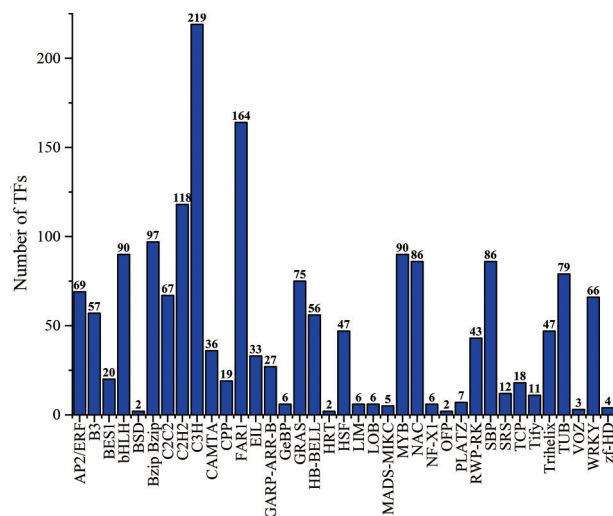


Figure 5 Classification and statistics of transcription factor annotations in *P. kingianum*. TFs: Transcription factors

测到 33 709 个 SSR 位点, 所有碱基重复类型中, 单碱基重复最多 (p1: 10 257); 最少的是五碱基重复 (p5: 148), 仅为单碱基重复的 0.14%, 可见不同类型的核

苷酸在数量上差距较大。其他 SSR 类型二碱基重复 (p2)、三碱基重复 (p3)、四碱基重复 (p4)、六碱基重复 (p6)、复合型重复 (s) 和 SSR 之间有重叠的复合型 SSR (c\*) 的数量分别为 4 815、5 952、259、373、1 707 和 10 180 (表 2)。

**Table 2** SSR site analysis of *P. kingianum*. SSR: Simple sequence repeats; c: Composite SSR; c\*: Composite type with overlap between SSRs; p1: Single base repeat; p2: Dibasic repeat; p3: Triple base repeat; p4: Tetra nucleotide duplication; p5: Pent nucleotide duplication; p6: Six base repeat

SSR type	c	c*	p1	p2	p3	p4	p5	p6
Number	1 707	10 180	10 275	4 815	5 952	259	148	373
Percent/%	5.06	30.2	30.48	14.28	17.66	0.77	0.44	1.11

**Table 3** SSR primer design table

ID	Primer sequence	Prime TM/°C	Prime GC percent/%	Primer pair product size/bp	SSR type	SSR	SSR length /bp	SSR start	SSR end
<i>Polygonatum kingianum</i> _transcript_61128	F:CCAGGAGAATGGGTTGGTCC R:AGAGAGGAACACAACAGCGG	60.034 59.966	60 55	280	p1	(A)10	10	22	31
<i>Polygonatum kingianum</i> _transcript_33808	F:ATGTTCTCAGAGCGTCAA R:CATCCACTCAGGGCATTGGT	59.965 60.034	50 55	103	c*	(CT)12(CTCT)6*	24	149	172
<i>Polygonatum kingianum</i> _transcript_52353	F:GGCGGGGAAGGGTTAAGAAA R:CAACGGAGTGCCTGGATTCT	59.961 60.036	55 55	200	p1	(T)10	10	1 952	1 961
<i>Polygonatum kingianum</i> _transcript_47350	F:GAAACAGGCGGTGACACTA R:CTGCAAGGAGAGATGCACCA	59.966 60.036	55 55	197	p3	(AGC)5	15	1 811	1 825
<i>Polygonatum kingianum</i> _transcript_21981	F:AGGCTGCAAATTCCTCCTCC R:TTTTGGAGGCGCAGGTAAT	60.034 59.961	55 50	239	p1	(T)10	10	16	25
<i>Polygonatum kingianum</i> _transcript_59587	F:GAGTACCCTTTTGCCGAGCT R:CTTCTGATCCGCTGCGATCT	60.036 59.968	55 55	279	c*	(T)13(TT)6*	12	1 332	1 343
<i>Polygonatum kingianum</i> _transcript_7136	F:AACACCATATCCTGCGTGCA R:GCATTGCAGGCCTCACAAAA	60.036 59.966	50 50	256	c	(T)10atcttttcc(GA)6	32	93	124
<i>Polygonatum kingianum</i> _transcript_2968	F:CTTCCCTCTGTCTGCAGCAA R:TAAGAGCAGGTCTCCTCCCC	59.964 60.032	55 60	227	p1	(T)10	10	4 338	4 347
<i>Polygonatum kingianum</i> _transcript_43887	F:CTGAGTTCCTTGCTGAGGG R:GAGAGCGGATATGGCGTTGA	60.036 59.968	60 55	243	p2	(CT)6	12	17	28
<i>Polygonatum kingianum</i> _transcript_58976	F:CCTCTACGTCAGGACCCAGA R:TGTTTCAGCGCCTTCAGCATA	60.034 60.036	60 50	143	p1	(T)11	11	63	73
<i>Polygonatum kingianum</i> _transcript_55221	F:GATTCTGAGGCAGCCCACTT R:CCCGGAAAGATTGTCCACA	60.035 59.962	55 55	170	p3	(GAA)8	24	306	329
<i>Polygonatum kingianum</i> _transcript_9971	F:CTGGATGGATTGGTGGGCTT R:ACGAATTGTCTCCGGCTTGT	60.033 59.965	55 50	203	p1	(T)10	10	590	599
<i>Polygonatum kingianum</i> _transcript_57199	F:CTCAAGGGCACTGTGACGAT R:AACTCAGATGTCTGCAGGGC	60.037 60.036	55 55	108	p3	(TTC)5	15	1 498	1 512
<i>Polygonatum kingianum</i> _transcript_9908	F:GAGCTACCATCACCTCGCTC R:CGACCAACAATGCAGCCAAA	59.969 59.968	60 50	127	p2	(CA)6	12	978	989
<i>Polygonatum kingianum</i> _transcript_28150	F:TTTTTCCCCCTCTCATCGGC R:AGAATTCGCCTTGCCATCCA	60.034 60.034	55 50	191	p2	(CT)6	12	20	31
<i>Polygonatum kingianum</i> _transcript_60339	F:CGTGGACGAGGATGTGGAAA R:ATTATGCCCCATGCCACAA	60.038 60.031	55 50	151	c*	(GT)9(GAGA)5 (GA)11*	40	14	53
<i>Polygonatum kingianum</i> _transcript_56385	F:AGTGGAGGTGAAGACCCTGT R:TGCCTTCATCCAATGCCAT	60.105 60.032	55 50	275	p1	(T)10	10	10	19
<i>Polygonatum kingianum</i> _transcript_15896	F:GTCTGTGCTGCAACACTGTG R:GCTCGTGAAGTGACTCCA	59.971 59.968	55 55	117	p3	(GAG)5	15	458	472
<i>Polygonatum kingianum</i> _transcript_56614	F:CCATCGAGAACCCTCATCG R:GAAGGGAGCCTGTGCGTTGA	59.967 59.965	60 55	146	c*	(TCTC)7(TCTCTC) 5*(TC)15*	30	101	130
<i>Polygonatum kingianum</i> _transcript_63298	F:TGTGATCGCTCTGTGCATGT R:ACAAGGACGACGAAGGCAAT	60.037 59.965	50 50	300	p1	(A)11	11	1 081	1 091
<i>Polygonatum kingianum</i> _transcript_44544	F:CTGCTGCTGCGATGGAATTC R:ATCATCAACACCTGGCTCCG	59.971 60.108	55 55	194	c	(A)10(GCT)9	106	3	108

**5.2 SSR 引物设计** 为进一步筛选可以在试验中应用的滇黄精 SSR 分子标记, 采用 Primer 3 对滇黄精转录本进行 SSR 引物设计, 共设计引物 24 034 对, 统计了其中 17 条与滇黄精多糖合成有关的转录本, 结果如表 3 所示。

## 6 多糖代谢酶基因分析

黄精多糖的生物合成是一个较为复杂的过程,首先卡尔文循环的产物3-磷酸甘油醛(GAP)在果糖二磷酸醛缩酶(ALDO)、果糖-1,6-二磷酸酶(FBP)、蔗糖磷酸合成酶(SPS)、蔗糖-6-磷酸酶(SPP)蔗糖合成酶(SUS)、磷酸葡萄糖变位酶(PGM)等酶的作用下生成蔗糖<sup>[21]</sup>;蔗糖在INV、PGM、尿苷二磷酸葡萄糖焦磷酸化酶(galU)的作用下生成UDP-葡萄糖,成为组成黄精多糖的第一类单糖;蔗糖在HK、果糖激酶(scrK)、甘露糖-6-果糖激酶(MPI)、磷酸甘露糖突变酶(PMM)和GMPP的作用下生成的GDP-甘露糖是组成黄精多糖的第二类单糖;GDP-甘露糖在甘露糖4,6-脱水酶(GMDS)、甘露糖-L-岩藻糖合酶的作用下生成的GDP-岩藻糖是组成黄精多糖的第3种单糖;UDP-葡萄糖在尿苷二磷酸葡萄糖-4,6脱水酶(RHM)和3,5差向异构-4-还原酶(UER1)的作用下生成UDP-鼠李糖,构成黄精多糖的第4种单糖,并且在尿苷二磷酸葡萄糖-4-表异构酶(GALE)和尿苷二磷酸-D-半乳糖脱氢酶(UGD)的作用下生成UDP-半乳糖,构成黄精多糖的第5种单糖;UDP-半乳糖在尿苷二磷酸-4-差向异构酶(UGE)、尿苷二磷酸芹糖/木糖合酶(AXS)、尿苷二磷酸阿拉伯糖-4-差向异构酶(UXE)的进一步作用下生成的UDP-D-吡喃木糖和UDP-L-阿拉伯糖是构成黄精多糖的两种多糖。这些单糖最后在糖基转移酶的作用下形成黄精多糖。本实验中,通过KEGG代谢通路富集分析,共发现与滇黄精多糖合成相关的19个基因或基因家族的转录本348个,其中糖酵解/糖异生代谢通路的转录本有127条,淀粉和蔗糖代谢通路的转录本有

144条,氨基酸糖和核苷酸糖代谢通路的转录本有85条,果糖和甘露糖代谢通路的转录本有69条。在这些基因家族或基因中,转录本数量最多的基因家族是SUS(62条),其次是INV(48条),最少的是FBP和PMM,只有1条(表4)。

INV将蔗糖转化为葡萄糖和果糖,是植物合成多糖的关键酶之一。将转录组测序得到的48个INV转录本表达量与滇黄精多糖含量进行相关性分析,其中INV12、INV23、INV25、INV36、INV44、INV46与滇黄精多糖含量具有显著的正相关,因此推测其可能在滇黄精多糖合成过程中起重要作用(图6)。

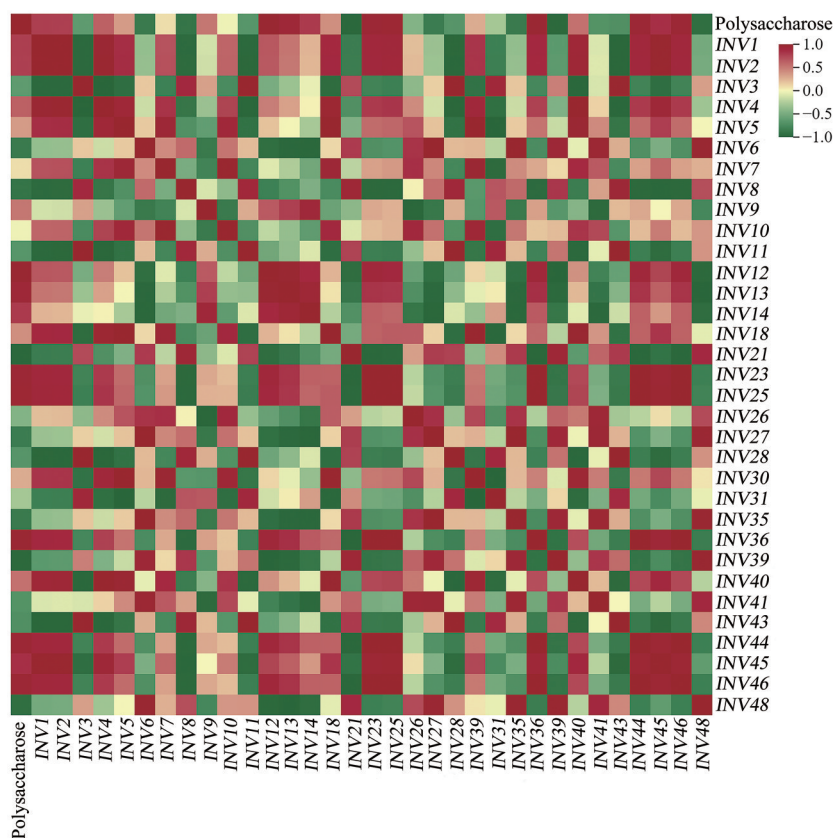
## 讨论

目前人们对于滇黄精的基因测序研究较少,限制了滇黄精的进一步开发利用,本研究通过PacBio测序平台首次对滇黄精的根茎进行全长转录组进行测序和分析,去除冗余后共获得55 454个转录本,其完整性和质量(N50, 3 212)远高于二代(N50, 1 260<sup>[22]</sup>; N50, 1 350<sup>[23]</sup>),可以为及后续滇黄精基因资源挖掘提供基础。通过NR数据库对比发现,有94.40%的基因得到注释,其中与滇黄精基因配比率最高的为天门冬科植物石刁柏(70.17%),此外共注释到黄精属植物18个,但其所占比例仅为0.41%,表明黄精属植物的已有基因资源相对匮乏。GO注释和KEGG注释结果显示滇黄精的转录组功能主要集中于生物学过程与代谢过程。这些基因注释和分类信息将有助于对滇黄精基因功能的理解与深入研究。

SSR分子标记技术是常用的遗传标记方法,具有重

**Table 4** Unigene related to the synthesis of polysaccharides from *P. kingianum*

Gene	No. of transcript	Enzyme	KEGG pathway	Enzyme code
<i>PGM</i>	22	Phosphoglucomutase	ko00010, ko00520	5.4.2.2
<i>ALDO</i>	10	Fructose-bisphosphate aldolase	ko00010	4.1.2.13
<i>GPI</i>	39	Glucose-6-phosphate isomerase	ko00010	5.3.1.9
<i>HK</i>	19	Hexokinase	ko00010, ko00051	2.7.1.1
<i>scrK</i>	36	Fructokinase	ko00010, ko00051	2.7.1.90
<i>FBP</i>	1	Fructose-1,6-bisphosphatase	ko00010	3.1.3.11
<i>MPI</i>	3	Mannose-6-phosphate isomerase	ko00051	5.3.1.8
<i>PMM</i>	1	Phosphomannomutase	ko00051	5.4.2.8
<i>GMPP</i>	5	Mannose-1-phosphate guanylyltransferase	ko00051	2.7.7.13
<i>GMDs</i>	5	Gdpmannose 4,6-dehydratase	ko00051	4.2.1.47
<i>RHM</i>	12	UDP-glucose 4,6-dehydratase	ko00520	4.2.1.76
<i>UXE</i>	4	UDP-arabinose 4-epimerase	ko00520	5.1.3.5
<i>AXS</i>	13	UDP-apiose/xylose synthase	ko00520	-
<i>UGE</i>	10	UDP-glucuronate 4-epimerase	ko00520	5.1.3.6
<i>GALE</i>	10	UDP-glucose 4-epimerase	ko00520	5.1.3.2
<i>UGDH</i>	14	UDP-glucose 6-dehydrogenase	ko00520	1.1.1.22
<i>SPS</i>	31	Sucrose-phosphate synthase	ko00500	2.4.1.14
<i>SUS</i>	62	Sucrose synthase	ko00500	2.4.1.13
<i>SPP</i>	3	Sucrose-6-phosphatase	ko00500	3.1.3.24
<i>INV</i>	48	Beta-fructofuranosidase	ko00500	3.2.1.26



**Figure 6** Correlation analyses between *INV* transcripts and polysaccharide content. *INV*: Beta-fructofuranosidase

复性好、稳定性强等优点, 广泛用于植物种质资源评价、遗传多样性分析等领域<sup>[24]</sup>。Song等<sup>[25]</sup>利用18对SSR分子标记, 对来自山西及周边省份的62个柴胡栽培和野生居群共619个植株进行遗传多样性以及居群遗传结构分析, 为柴胡的种质资源利用、遗传变异研究以及柴胡优质种质资源开发提供理论依据。本文在滇黄精转录本中共检测到33 709个SSR位点, 单核苷酸重复到六核苷酸重复以及复合型重复和SSR之间有重叠的复合型SSR均有检测到, 其重复类型主要单碱基重复和SSR之间有重叠的复合型SSR为主, 占总SSR位点的60.68%。利用Primer 3.0软件对获得的滇黄精SSR进行引物设计, 筛选得到了24 034对SSR引物, 显示了滇黄精SSR位点的多态性。本文统计了其中与滇黄精多糖合成相关酶的转录本的SSR引物设计信息, 以期为滇黄精SSR分析以及后续的多糖合成奠定理论基础。

作为药食同源植物, 滇黄精具有十分广阔的发展前景。中国药典规定多糖作为滇黄精唯一质控标准, 因此探究滇黄精中多糖类成分的生物合成途径对于滇黄精的进一步开发利用具有重要意义。本文利用第三代转录组测序技术获得的非冗余转录本中共初步挖掘得到参与滇黄精多糖生物合成的19个基因或基因家族的348条转录本, 远远超过Wang等<sup>[10]</sup>通过二代转录

组测序获得的参与滇黄精多糖生物合成的相关基因或基因家族数量, 说明本实验中使用全长转录组测序测得的转录本更加准确。在得到的19个相关基因中, 除了*FBP*和*PMM*两个基因只检测到1条转录本外, 其余基因均得到多条转录本, 其中*SUS*转录本高达62条, 这为后续多糖生物合成关键酶基因的挖掘提供了更多的候选基因。

随着转录组测序技术的不断发展以及数据库的不断完善。将结果与数据库进行比对, 可得到更多的滇黄精遗传信息及有助于滇黄精多糖合成途径的揭秘。本研究首次利用PacBio Sequel平台获得了滇黄精全长转录组, 分析了滇黄精多糖类成分的生物合成途径, 为滇黄精多糖生物合成的分子机制奠定了基础。

**作者贡献:** 米琪与赵艳丽共同完成实验内容及论文撰写工作; 徐萍负责实验样品采集以及论文撰写工作; 于梦雯、张萱和涂振华协助完成部分实验内容; 李春华为本实验提供了实验样品; 陈佳指导实验和平台支撑; 郑国伟设计方案并指导论文的撰写和修改。

**利益冲突:** 所有作者均声明不存在利益冲突。

## References

- [1] Chinese Pharmacopoeia Commission. Pharmacopoeia of the

- People's Republic of China (中华人民共和国药典) [S]. Vol 1. 2020 Ed. Beijing: China Medical Science Press, 2020: 319-320.
- [2] Yin SY, Qian HL. Research progress in diversity of origin plants of Chinese medicinal *Polygonati Rhizoma* and its utilization [J]. Chin Wild Plant Resour (中国野生植物资源), 2022, 41: 49-57.
- [3] Zhao WL, Zhao Y, Tseng Y. Research progress on pharmacological effects of *Polygonati Rhizoma* [J]. China Tradit Herb Drugs (中草药), 2018, 49: 4439-4445.
- [4] Zhang K, Zhang Z, Fan YF, et al. Nutritional and bioactive components of *Polygonati Rhizoma* and *Polygonati Odorati Rhizoma* with medicinal and edible property [J]. Mod Chin Med (中国现代中药), 2022, 24: 1463-1472.
- [5] Sun SY, Wang HL, Chen SS, et al. Comparative study on pharmacological and edible components of different strains of *Polygonatum polyflora* [J]. J Food Saf Qual (食品安全质量检测学报), 2022, 13: 7272-7278.
- [6] Zhou ZG, Gong MH, Wang XH, et al. Research progress on pharmacological action and mechanism of *Polygonatum sibiricum* polysaccharide [J]. J Liaoning Univ Tradit Chin Med (辽宁中医药大学学报), 2021, 23: 1-4.
- [7] Liang PG, Jian S, Jian Y, et al. Response surface optimization of extraction and structure and activity analysis of polysaccharides from *Polygonatum kingianum* Coll. et Hemsl [J]. J South Agric (南方农业学报), 2021, 52: 3434-3445.
- [8] Xu HL, Mwang J, Fan SM, et al. Metabolic pathway analysis of *Polygonatum cyrtoneuma* and *Polygonatum filipes* based on transcriptomic data [J]. J Chin Med Mater (中药材), 2020, 43: 2663-2668.
- [9] Wang SQ, Wang B, Hua WP, et al. *De novo* assembly and analysis of *Polygonatum sibiricum* transcriptome and identification of genes involved in polysaccharide biosynthesis [J]. Int J Mol Sci, 2017, 18: 1905.
- [10] Wang Y, Tan YF, Zhang LP, et al. Characteristic analysis of *Polygonatum sibiricum* transcriptome and study of related functional genes [J]. J Sichuan Norm Univ Nat Sci (四川师范大学学报 自然科学版), 2022, 45: 103-109.
- [11] Zhang D. Comparative Transcriptomics of *Panax notoginseng* in Different Cultivation Modes Provides Insights into Ginsenoside Biosynthesis (不同耕作模式下三七人参皂苷生物合成的比较转录组学研究) [D]. Kunming: Kunming University of Science and Technology, 2017.
- [12] Rhoads A, Au KF. PacBio sequencing and its applications [J]. Genom Proteom Bioinf, 2015, 13: 278-289.
- [13] Zhang ZJ, Liu YR, Zhang SJ, et al. The principle of the third generation sequencing technology and its application in biological field [J]. Chin J Anim Sci (中国畜牧杂志), 2020, 56: 11-15.
- [14] Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND [J]. Nat Methods, 2015, 12: 59-60.
- [15] Sun L, Luo H, Bu D, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts [J]. Nucleic Acids Res, 2013, 41: e166.
- [16] Kang YJ, Yang DC, Kong L, et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features [J]. Nucleic Acids Res, 2017, 45: W12-W16.
- [17] Li A, Zhang J, Zhou Z. PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme [J]. BMC Bioinformatics, 2014, 15: 311.
- [18] Zheng Y, Jiao C, Sun H, et al. ITAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases [J]. Mol Plant, 2016, 9: 1667-1670.
- [19] 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: 3411-3422.
- [20] Hu RL, Yuan C, Niu Y, et al. Regulation of plant MYB transcription factors in anther development [J]. Chin J Biotechnol (生物工程学报), 2020, 36: 2277-2286.
- [21] Chen L, Xu S, Liu Y, et al. Identification of key gene networks controlling polysaccharide accumulation in different tissues of *Polygonatum cyrtoneuma* Hua by integrating metabolic phenotypes and gene expression profiles [J]. Front Plant Sci, 2022, 13: 1012231.
- [22] Tao P, Liu Y, Tang ZW, et al. Transcriptome-based screening and validation of key enzyme genes for *Polygonatum* polysaccharide metabolism [J]. Chin J Exp Tradit Med Form (中国实验方剂学杂志), 2023, 29: 157-167.
- [23] Xiao YZ, Han SM, Qin Z, et al. Analysis of transcriptome sequencing and related genes of flavonoid biosynthesis from *Polygonatum kingianum* [J]. J Henan Agric Univ (河南农业大学学报), 2020, 54: 931-940.
- [24] Li S, Kou PW, Xu YK, et al. Development and application of SSR markers of *Saposhnikovia divaricata* based on transcriptome [J]. China J Chin Mater Med (中国中药杂志), 2024, 49: 717-727.
- [25] Song Y, Zhang XR, Li Z, et al. Genetic diversity and genetic structure analysis of *Bupleurum chinense* DC. based on SSR molecular marker [J]. Acta Pharm Sin (药学报), 2022, 57: 1193-1202.